



ESCUELA TECNICA SUPERIOR DE INGENIEROS INDUSTRIALES Y DE TELECOMUNICACIONES

Titulación:

**INGENIERO TECNICO DE TELECOMUNICACIONES,
ESPECIALIDAD IMAGEN Y SONIDO**

Título del proyecto:

ESTUDIO DE PITCH EN SEÑALES POLIFÓNICAS

**Raúl Zalba Oteiza
Miroslav Zivanovic**

Pamplona, 17 de Noviembre de 2009

INDICE

1. INTRODUCCION	4
2. MOTIVACION	6
2.1 Estado del arte	6
2.1.1 Representaciones tiempo-frecuencia lineales: STFT	7
2.1.2 Espectrograma	8
2.2 Problemática	9
3. OBJETIVOS	13
4. CARACTERIZACION DE SONIDOS	13
4.1 Frecuencia fundamental	13
4.2 Clasificación de sonidos	15
4.2.1 Número de canales	15
4.2.2 Cuasi- estacionarios y no estacionarios	15
4.2.3 Monofónicos y polifónicos	18
4.2.4 Mono tímbrico y multitímbrico	19
4.3 Transcripción automática musical	19
4.4 Estimación multi-pitch	20
5. ANALISIS Y SINTESIS DE AUDIO	22
5.1 Esquema general	22
5.2 Esquema particular	24
6. FUNCIONAMIENTO DE LA HERRAMIENTA	27
6.1 Calculo de espectrograma	27
6.2 Clasificación de picos espectrales	28
7. RESULTADO EXPERIMENTAL	31
7.1 Procedimiento	31

7.2 Cálculo de la frecuencia fundamental a través del descriptor:.....	32
7.3 Cálculo de la frecuencia fundamental a través del método de autocorrelación.....	39
7.4 Comparación de métodos para señal con una sola fuente.....	41
7.5 Cálculo de la frecuencia fundamental a través del método del descriptor para una señal con dos fuentes:.....	42
7.6 Cálculo de la frecuencia fundamental a través del método de autocorrelación para una señal con dos fuentes:.....	50
7.7 Comparación de Métodos para señal con dos fuentes:.....	51
 8. CONCLUSIONES Y LINEAS FUTURAS.....	52
8.1 Conclusiones.....	52
8.2 Líneas Futuras.....	53
 9. BIBLIOGRAFIA.....	54

1. INTRODUCCION:

El sistema auditivo humano posee la capacidad de procesar e identificar eficientemente los sonidos que están activos a su alrededor, incluso si dichos sonidos proceden de diferentes fuentes sonoras y/o suceden de manera simultánea.

A pesar de los grandes avances y de los resultados prometedores obtenidos en la última década, hoy por hoy se puede decir que el mejor sistema de transcripción musical polifónico existente es el propio sistema auditivo humano. Como consecuencia, este campo de investigación está totalmente abierto ya que la alta tasa de error proporcionada por los sistemas actuales de transcripción hace inviable el desarrollo de una herramienta práctica para señales musicales polifónicas del mundo real.

Hoy en día, las principales limitaciones que presentan las señales musicales polifónicas para un sistema de transcripción vienen determinadas por el número de fuentes sonoras concurrentes, las relaciones existentes entre las frecuencias fundamentales de los eventos musicales interpretados por dichas fuentes y las interferencias ocasionadas por los sonidos. Los problemas surgen ante la existencia de sonidos concurrentes cuyas frecuencias fundamentales presentan relaciones armónicas entre sí. Esta situación es muy común en la música Occidental y aumenta enormemente la dificultad a la hora de la extracción del pitch ya que cada evento musical lleva implícito la distorsión del resto de eventos que presentan solapamiento en frecuencia.

El método de síntesis considerado para la realización de este proyecto es el de síntesis aditiva, que modela un sonido como un conjunto de sinusoides estables, llamadas parciales y un ruido o componente residual. Esta técnica está construida sobre la base del análisis de Fourier, que indica que cualquier forma de onda periódica puede ser modelada como una suma de sinusoides de diferentes amplitudes y fases y frecuencias armónicas. La misma separación puede ser aplicada a sonidos vocales e incluso a sonidos no musicales encontrados en la naturaleza.

Los sonidos producidos por un instrumento musical, pueden ser modelados como la suma de un conjunto de sinusoides más un ruido residual. La parte sinusoidal o determinista, se corresponde con los principales modos de vibración del sistema, mientras que la parte residual o el ruido comprende la energía producida por el mecanismo de excitación que no es transformado por el sistema a vibraciones estacionarias, más cualquier otro componente de energía que no es sinusoidal en la naturaleza.

La señal determinista es una señal predecible a partir de mediciones sobre un intervalo continuo, por tanto llamaremos señal determinista a la suma de componentes casi sinusoidales o sinusoidales cuya amplitud y frecuencia varia a través del tiempo, cada senoide modela una estrecha banda del sonido original y se describe mediante una función de amplitud y una función de frecuencia. La señal estocástica queda totalmente descrita por su densidad espectral de potencia, que proporciona la potencia esperada para cada banda frecuencial. Cuando una señal se considera estocástica, no es necesario conservar información sobre su fase instantánea o su magnitud exacta para cada FFT individual.

2. MOTIVACION

2.1 Estado del arte

Los primeros avances en el campo de las transformadas tiempo frecuencia se produjeron con motivo de estudios teóricos relacionados con el intento de asentar las bases matemáticas que permitieran la obtención de funciones capaces de representar simultáneamente el hecho físico de la variación del espectro con el tiempo, y poder utilizar esta función de una variable tal y como se emplea la función de densidad de energía de una variable. De esta forma, se podría tener una señal completamente identificada por las dos propiedades básicas que la caracterizan: su amplitud en el tiempo en el que ocurre y la amplitud de cada frecuencia o frecuencias que contiene.

Como primera aproximación a las transformadas tiempo-frecuencia, y empleada anteriormente el desarrollo de las representaciones tiempo-frecuencia cuadráticas, una posibilidad consiste en analizar la señal con transformadas de Fourier consecutivas y de corta duración STFT. Así tomando pequeñas variaciones de la señal se apreciara que la distribución espectral de cada una de ellas varia con el tiempo, pudiendo conocer con más precisión la evolución de las componentes espectrales. La equivalencia directa dentro de las representaciones cuadráticas es el espectrograma.

En los últimos años se han desarrollado diferentes representaciones de tiempo-frecuencia puesto que dependiendo de la aplicación resulta conveniente que se satisfagan diferentes propiedades, o tengan características de resolución y términos de interferencia concretos. Por un lado tenemos las representaciones tiempo-frecuencia lineales y por otro lado las cuadráticas o bilineales que se pueden agrupar dentro de las llamadas clase de Cohen. Se hará una breve descripción general de cada uno de los tipos más importantes. Como diferencia principal entre las representaciones tiempo – frecuencia lineales y las cuadráticas, se puede decir que mientras en las primeras se muestra la señal descompuesta en el plano t-f basados en la amplitud de la señal temporal, en las segundas, la descomposición se realiza basándonos en la energía, y lo que se distribuye en el plano t-f es la energía de la señal, y dado que la energía de la señal requiere de la señal al cuadrado para obtenerse, se dice que estas representaciones son cuadráticas.

2.1.1. Representaciones tiempo-frecuencia lineales: STFT

Este tipo de representaciones son ampliamente utilizadas debido a su sencillez y similitud con los métodos espectrales clásicos en cuanto a concepto y estrategia de cálculo. Reciben este nombre porque en su obtención, la señal a analizar no se multiplica por sí misma como ocurre en las representaciones cuadráticas y además cumplen el principio de superposición. También se les llama descomposiciones atómicas, ya que proyectan la señal en el plano de forma que se obtienen celdas que conforman la representación global.

La transformada corta de Fourier (STFT) es muy utilizada en diferentes aplicaciones debido a su sencillez de implementación, pero dada su baja resolución espectral y temporal se tiende a dejar de utilizarla a favor de las representaciones bilineales o cuadráticas.

Si para una señal de larga duración se toman pequeños tramos de ella realizando un enventanado, y se calcula la transformada de Fourier de cada tramo podremos conocer las componentes frecuenciales en una ventana de tiempo pequeña. Así analizando cada transformada de Fourier se observa el cambio de dichas componentes frecuenciales. La STFT se define como:

$$\text{STFT}(t, f) = \int_T [X(T) \cdot Y(T - t)] \cdot e^{-j2\pi fT} \cdot dT$$

La elección del tamaño de la ventana $Y(t)$ es crítica a la hora de obtener una buena resolución. Si lo que se desea es obtener una resolución grande en el dominio temporal elegiremos una ventana de poca longitud, en cambio, la distribución espectral estará muy dispersa y no podremos conocer con exactitud sus componentes. Por contrario, la elección de una ventana grande nos mostrara con más precisión las componentes en frecuencia, pero perderemos idea del momento en que se produjeron cada una de ellas.

2.1.2. Espectrograma:

El espectrograma es el modulo al cuadrado de la STFT. La STFT es en general compleja y depende de la ventana de análisis, que suele elegirse suave con un máximo en el origen y decreciente hacia $\pm\infty$. También suele ser real y par.

La discretización de la STFT es sencilla. En el tiempo, para evitar el efecto del solapamiento espectral hace falta que la suma de anchos de banda de la señal y de la ventana no sobrepasen la mitad de la frecuencia de muestreo. Por otro lado, esto no se cumple estrictamente nunca, porque las ventanas utilizadas son de duración finita.

La principal ventaja del espectrograma es que sus términos cruzados son poco apreciables. Únicamente aparecen allí donde hay señal, con lo que quedan prácticamente enmascarados por la señal.

2.2 Problemática

La problemática con la que se encuentra al realizar este proyecto se puede dividir en dos casos diferentes:

- Trama poco estacionaria:

Se conoce que para una señal estacionaria o cuasi-estacionaria su espectro tiene la característica de que las frecuencias que se obtienen están colocadas en frecuencias múltiplos de la frecuencia fundamental (F_0), con lo que estarán situadas en $F_0, 2F_0, 3F_0, 4F_0, \dots$

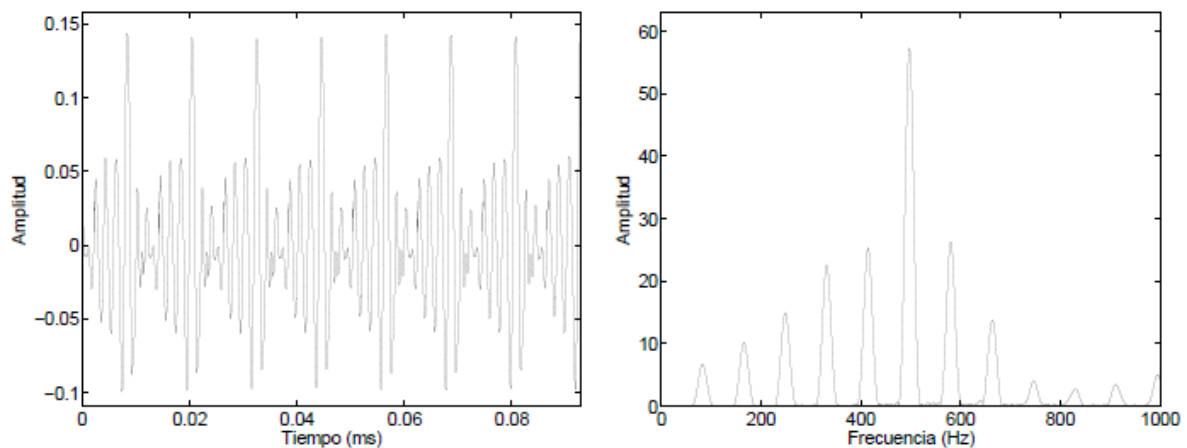


Figura 2.1: La figura de la izquierda muestra una señal cuasi-periódica en el dominio del tiempo, mientras que la figura de la derecha muestra su representación frecuencial.

Se observa como la frecuencia fundamental de esta señal está en torno a los 80 Hz, y los demás armónicos se sitúan a frecuencias múltiplos de esta, $2F_0$ (160 Hz), $3F_0$ (240 Hz), $4F_0$ (320 Hz),....

Con lo que se podría obtener fácilmente la frecuencia fundamental y sus diferentes armónicos. Pues bien en el caso de tener una señal que no sea estacionaria, esto no ocurrirá y los armónicos dejarán de estar a frecuencias múltiplo de la principal, lo que dificultará poder calcular la frecuencia fundamental así como los diferentes armónicos de la señal.

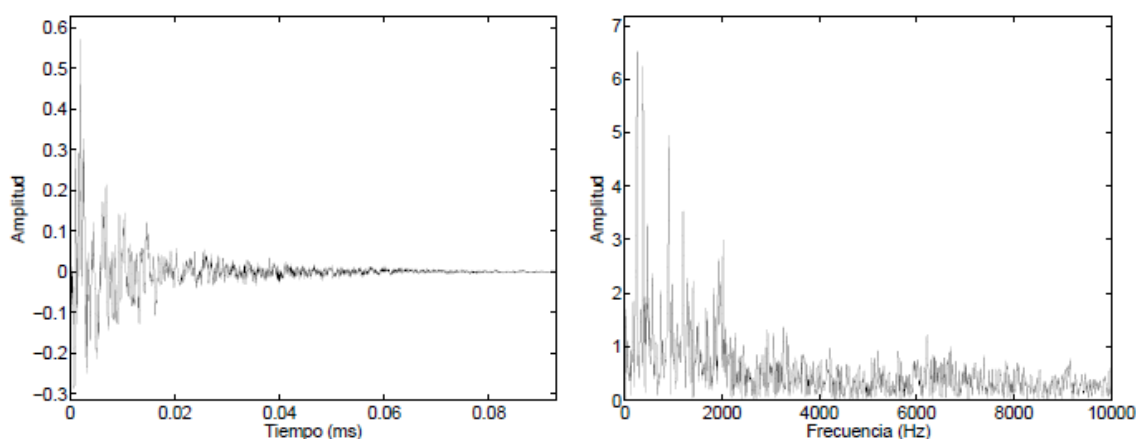


Figura 2.2: La figura de la izquierda muestra una señal aleatoria en el dominio del tiempo, mientras que la figura de la derecha muestra su representación frecuencial.

En este caso se observa como la señal no es nada estacionaria, y si nos fijamos en su representación frecuencial, podemos ver que las diferentes componentes frecuenciales aparecen aleatoriamente y no están colocadas a frecuencias múltiplo de la frecuencia fundamental. Con lo que sería más complicado poder sacar la frecuencia fundamental junto con sus armónicos.

- Solapamiento de frecuencias debido a varias fuentes musicales.

El mayor problema con el que se encuentra en sonidos con más de una fuente, es el solapamiento de frecuencias, lo que dificulta su separación. Se define intervalo musical a la diferencia en frecuencia entre dos eventos musicales concurrentes. Los intervalos musicales, a la vez que son una situación muy común en la música, son uno de los escenarios más complejos de analizar desde un punto de vista del procesado de señal ya que suelen presentar el problema del solapamiento de parciales, es

decir, sonidos donde algunos de sus parciales están solapados entre sí.

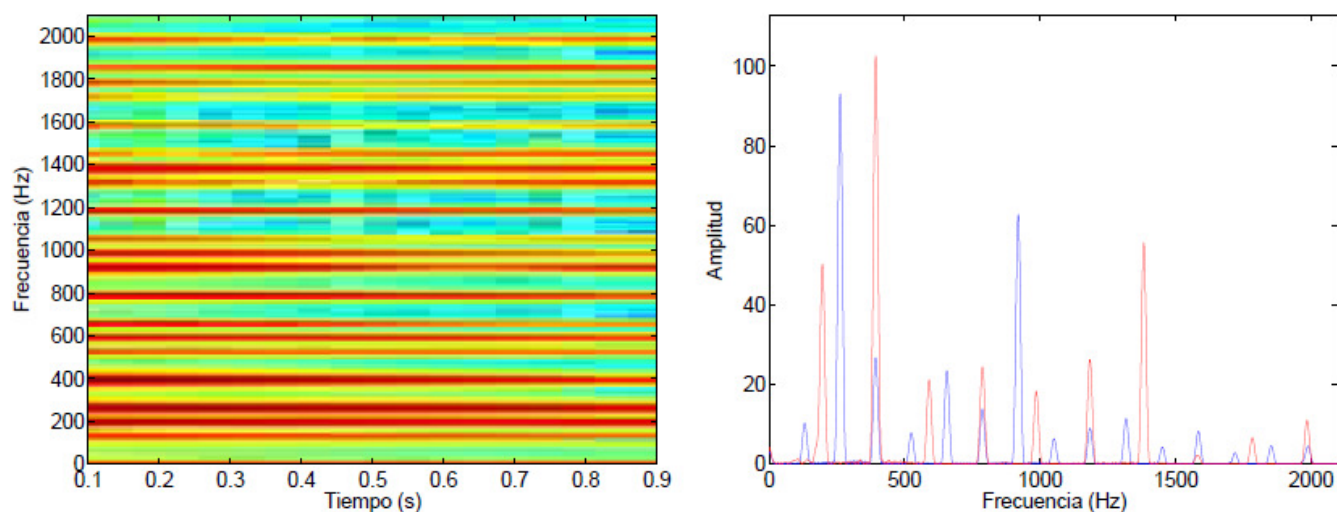


Figura 2.3: La figura de la izquierda es el espectrograma de una señal compuesta por 2 notas diferentes de un piano, la figura de la derecha es la representación frecuencial de dicha señal.

En la figura 2.3 de la derecha se observa el solapamiento de parciales, se representa en colores distintos, rojo y azul, cada una de las 2 notas, se puede comprobar cómo cada parcial múltiplo de 3 del evento de menor frecuencia se superpone a cada parcial múltiplo de 2 del evento de mayor frecuencia, con lo que será muy difícil poder recuperar cada una de ellas por separado.

3. OBJETIVOS

El objetivo principal de este proyecto es el cálculo de la frecuencia fundamental de una señal de audio polifónica. La frecuencia fundamental o el Pitch es un atributo perceptual que permite ordenar de manera creciente los sonidos en una escala de frecuencia logarítmica, es decir, representa la nota percibida de un evento musical. Un sonido presenta un determinado pitch, si la frecuencia de una senoide de amplitud arbitraria puede ser correspondida con el sonido percibido.

El concepto físico asociado al atributo pitch se denomina frecuencia fundamental (F_0), la cual está definida sólo para sonidos periódicos o cuasi-periódicos. Debido a que en la mayoría de las situaciones los términos pitch y frecuencia fundamental coinciden, en este proyecto ambos términos son considerados iguales.

Se van a utilizar dos técnicas diferentes para el cálculo del Pitch:

- Método del descriptor.
- Método de autocorrelación.

También se utilizarán dos señales diferentes:

- La primera constará de una sola fuente.
- La segunda constará de dos fuentes.

4. CARACTERIZACION DE SONIDOS

De cara al análisis de señales musicales de audio, existen varios atributos perceptuales subjetivos que resultan de interés a la hora de caracterizar los sonidos: frecuencia fundamental (F_0), intensidad sonora, duración, timbre, etc. Pero este proyecto se va a centra en la frecuencia fundamental.

4.1 Frecuencia Fundamental:

Una señal $x(t)$ es periódica si se repite cada cierto intervalo de tiempo. Se denomina periodo fundamental T_0 , en segundos, al mínimo intervalo de tiempo que satisface $x(t) = x(t + T_0)$; $\forall t$. No obstante, las señales periódicas existentes en el mundo que nos rodea son señales causales, reales y con una duración innata, lo que significa que si existe un periodo fundamental T_0 , entonces existen infinitos periodos (ver ecuación 2.4) para los cuales la señal $x(t)$ es periódica.

$$x(t) = x(t + nT_0); n \in (-\infty; \infty); n \in \mathbb{Z}$$

Se define la frecuencia fundamental F_0 , en Hertzios, de una señal periódica $x(t)$ como la inversa del periodo fundamental T_0

$$F_0 \text{ (Hz)} = \frac{1}{T_0 \text{ (s)}}$$

Las señales periódicas $x(t)$ se pueden descomponer mediante una suma de sinusoides, donde cada senoide $s_k(t)$ se define como la parte real de una exponencial compleja, la cual, podría ser representada mediante una función coseno con tres parámetros (amplitud $|a_k|$, frecuencia F_0 y fase ϕ_k).

$$s_k(t) = \text{Re} (a_k \cdot e^{jk2\pi F_0 t})$$

$$x(t) = \sum_{k=-\infty}^{\infty} s_k(t) = \sum_{k=-\infty}^{\infty} a_k \cdot e^{jk2\pi F_0 t}$$

Debido a que las exponenciales complejas son auto funciones de los sistemas lineales e invariantes en el tiempo (LTI), toda combinación lineal de sinusoides a la entrada de un sistema LTI proporciona a la salida la misma combinación lineal de sinusoides con un escalado en la amplitud compleja. Esta propiedad característica de las sinusoides permite que éste tipo de señales sean idóneas para modelar señales musicales ya que según el teorema de Fourier, cada señal periódica $x(t)$ se puede descomponer en infinitas sinusoides relacionadas armónicamente entre sí.

Sin embargo, debido a que las señales musicales existentes en el mundo son no estacionarias a largo plazo, es decir, sus propiedades estadísticas media μ_x y autocorrelación R_x son variantes con el tiempo, se consideran pequeños intervalos de tiempo denominadas tramas (del orden de decenas de milisegundos) dentro de los cuales la señal es estacionaria o cuasi-periódica. Así, para una señal cuasi-periódica $p(t)$, el teorema de Fourier señala que se puede descomponer en M sinusoides relacionadas armónicamente entre sí

En señales cuasi-periódicas, el concepto de armónico se reemplaza por parcial, el cual define componentes f_p cuyas frecuencias no se encuentran exactamente en múltiplos enteros de la frecuencia fundamental.

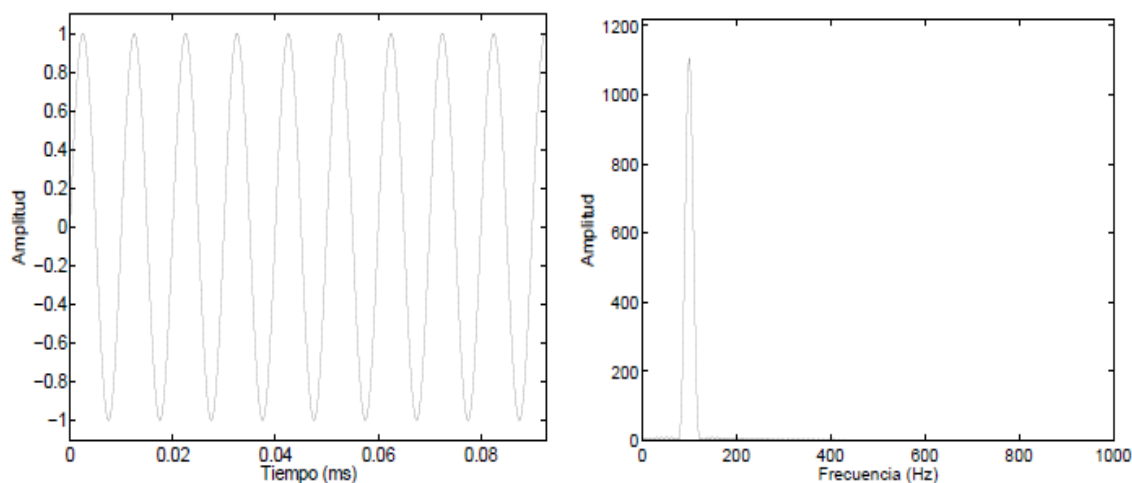


Figura 4.1: La figura de la izquierda muestra una señal periódica de un senoide en el dominio del tiempo con periodo $T=10$ ms. La figura de la derecha muestra la misma señal en el dominio de la frecuencia, aparece un solo pico que será la frecuencia fundamental, $F_0=100$ Hz.

4.2 Clasificación de sonidos:

Los diferentes sonidos que existen en el mundo real se pueden clasificar atendiendo a diversos criterios:

4.2.1 Numero de canales:

- Los sonidos monocanal son aquellos sonidos que han sido grabados utilizando un solo micrófono. Este tipo de sonidos son los más complicados de analizar desde un punto de vista del procesado de señal ya que no proporcionan ningún tipo de información espacial, es decir, la señal monocanal es la suma interferente de todas las fuentes sonoras activas.
- Los sonidos estéreo son aquellos sonidos que han sido grabados utilizando dos micrófonos independientes. La señal estéreo, compuesta por dos canales monocanal relacionados entre sí, proporciona información espacial acerca de la localización virtual de las fuentes sonoras. Este tipo de sonidos es el que soporta la mayor parte de la música comercial.
- Los sonidos multicanal son aquellos sonidos que han sido grabados utilizando tres o más micrófonos, es decir, están formadas por tres o más canales monocanales relacionados todos ellos entre sí donde cada altavoz trata de forma independiente un rango de frecuencias determinado.

4.2.2 Cuasi-estacionarios y no estacionarios:

- Un sonido cuasi-estacionario es aquel, que está compuesto por un conjunto de sinusoides cuyas frecuencias están relacionadas armónicamente entre sí. Esto implica que el espectro de un sonido armónico está formado por un conjunto de picos espectrales prácticamente igualmente espaciados cada F_0 Hz.

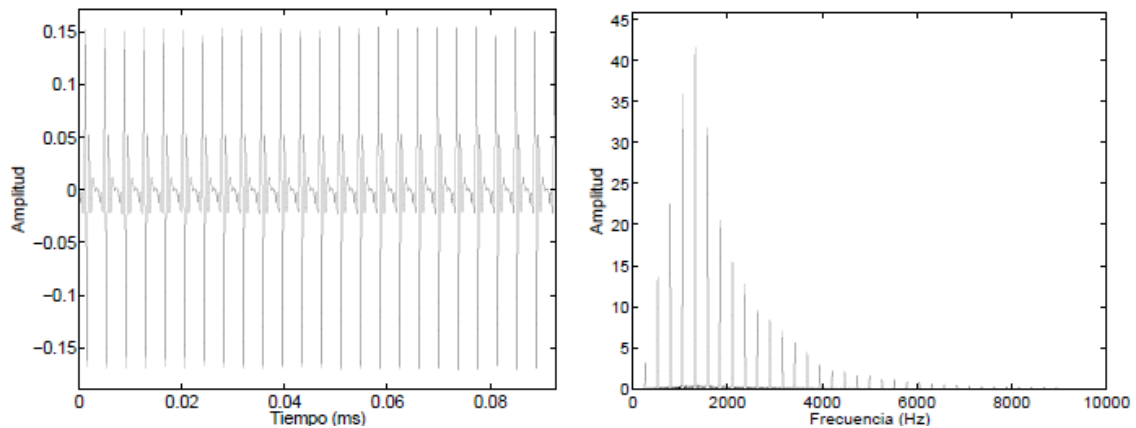


Figura 4.2: Señal cuasi periódica, la figura de la izquierda muestra la señal en el dominio del tiempo y la figura de la derecha muestra el espectro de la señal de la izquierda.

En la figura 4.2 se observa como los picos en el dominio de la frecuencia aparecen igualmente espaciados en múltiplos de la frecuencia fundamental.

- Sin embargo, muchos sonidos existentes en el mundo real son no estacionarios, el cual es el responsable de que se produzca una desviación en frecuencia de los parciales respecto a sus armónicos.

Por ejemplo en el inicio de una señal de voz o musical antes de que empiece la voz de una persona o el comienzo de un instrumento aparece cierto ruido y aunque inventanemos la señal con una ventana muy pequeña para que la trama sea la más cuasi estacionaria posible no siempre se consigue.

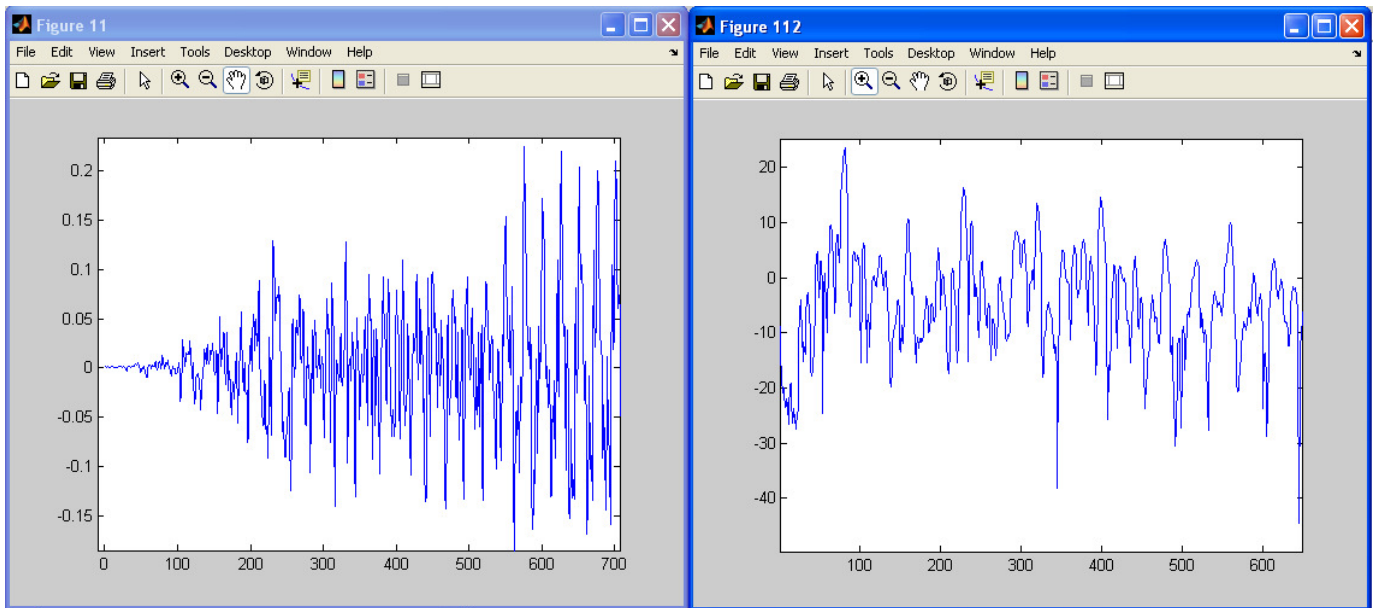


Figura 4.3: Esta señal es el comienzo de una nota musical de una guitarra, la figura de la izquierda muestra la señal en el dominio del tiempo y la de la derecha en el dominio de la frecuencia.

En el comienzo de la figura 4.3 se puede ver que la señal no es cuasi-periódica, por lo que al analizar su espectro obtenemos una serie de picos arbitrariamente colocados en diferentes frecuencias, tampoco están claramente definidos ya que algunos armónicos constan de dos picos, otros son muy anchos, etc.

Si se deja que esta misma señal avance en el dominio temporal, la señal se va haciendo cada vez más cuasi-periódica.

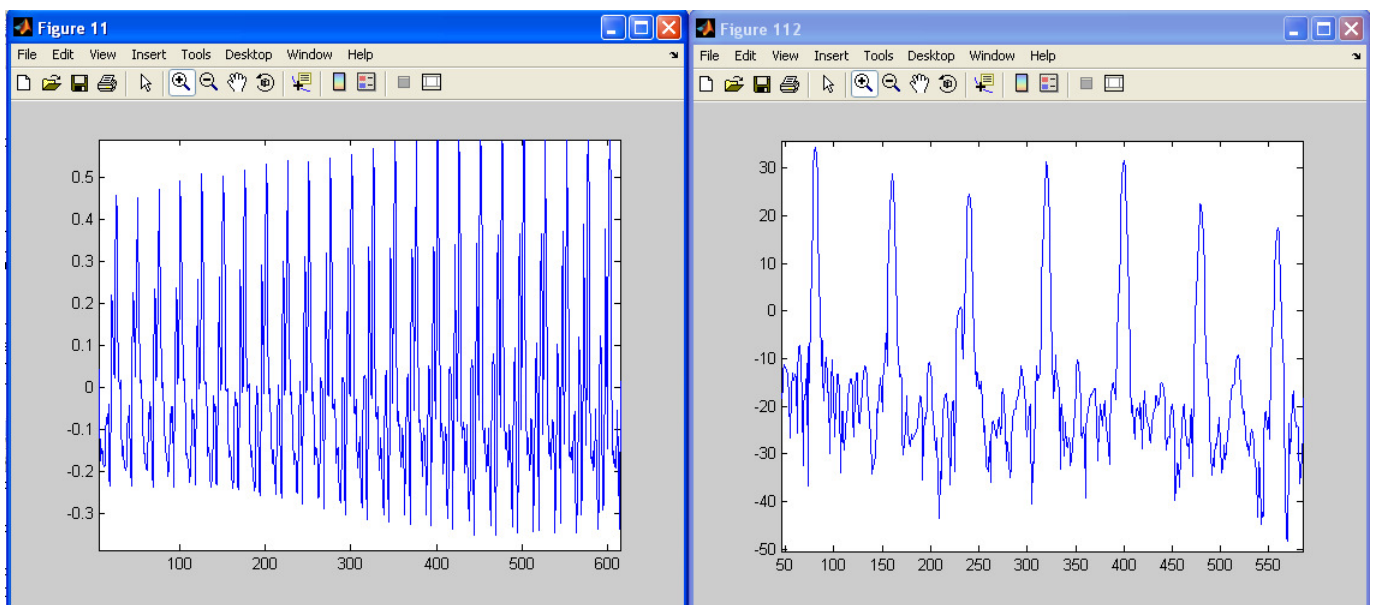


Figura 4.4: Esta señal es la misma que la 4.3 solo que unas tramas más adelante, la figura de la izquierda muestra la señal en el dominio del tiempo y la de la derecha en el dominio de la frecuencia.

Se puede observar que la figura 4.4 empieza a ser más cuasi-periódica en el dominio temporal, por lo que si se analiza la figura en el dominio frecuencial, se aprecian unos picos bien definidos y equidistantes, en múltiplos de la frecuencia fundamental.

4.2.3 Monofónicos y polifónicos:

- Se define sonido monofónico como aquel sonido en el cual sólo existe una frecuencia fundamental activa en cada instante temporal.
- Se define sonido polifónico como aquel sonido en el cual existe dos o más frecuencias fundamentales, procedentes de la misma o diferentes fuentes sonoras, activas cada instante temporal
- Al aumentar la polifonía aumenta la dificultad para distinguir qué pitches son los responsables del espectro analizado. Esta dificultad está íntimamente relacionada con el número variable de parciales que componen cada evento musical, solapamiento en frecuencia de diferentes parciales pertenecientes a diferentes eventos concurrentes, diferentes envolventes espectrales generadas por diferentes fuentes sonoras, etc.

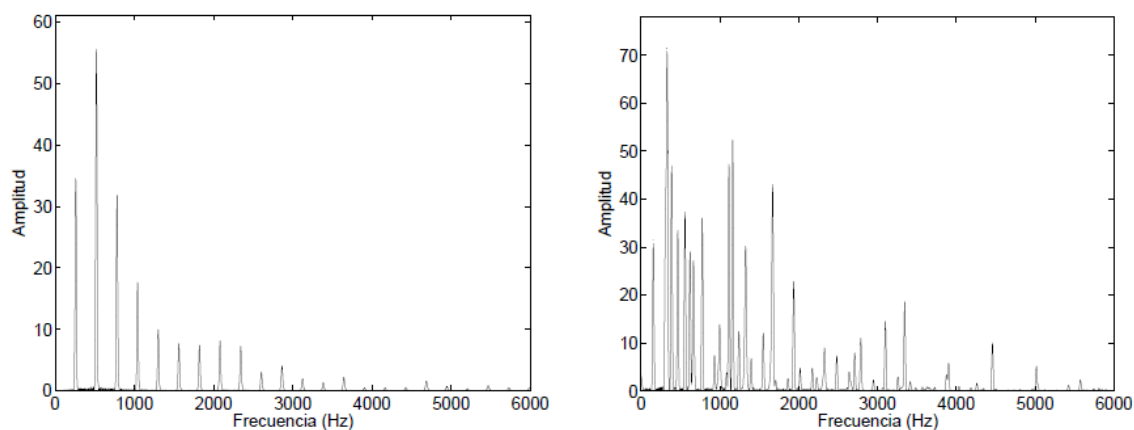


Figura 4.5: *La figura de la izquierda se corresponde con una señal monofónica que está formada por la nota de una flauta. La figura de la derecha se corresponde con una señal polifónica que está formada por tres instrumentos diferentes, una flauta, un clarinete y un oboe.*

En la figura de la derecha se observa la complejidad de separar correctamente la frecuencia fundamental de cada instrumento.

4.2.4 Monotímbrico y multitímbrico:

Atendiendo al número de instrumentos musicales diferentes, los sonidos polifónicos pueden ser:

- Se define un sonido monotímbrico como aquel sonido polifónico en el que todos los eventos han sido generados por el mismo instrumento musical, presentando el mismo tipo de envolvente espectral o timbre.
- Se define un sonido multitímbrico como aquel sonido en el que se han utilizado dos o más instrumentos musicales con timbre diferente para generar la señal polifónica.

4.3 Transcripción automática musical

Se define la transcripción musical como el proceso de analizar una señal de audio y poder extraer, mediante técnicas de procesamiento de señal, los parámetros (pitch, onset y duración) de los sonidos que componen dicha señal. La extracción de dichos parámetros se plasma de forma escrita utilizando una notación o representación simbólica (símbolos denominados notas o eventos musicales) que indican el pitch, onset y duración de cada evento existente.

4.4 Estimación multi-pitch

Las señales más adecuadas para evaluar técnicas de estimación multi-pitch son las señales musicales polifónicas. Considerando una señal polifónica, se define la estimación multi-pitch como el proceso de determinar el número de pitches activos en un instante temporal, proporcionando la amplitud y frecuencia pertenecientes a cada componente armónica de cada sonido determinado. Sin embargo, este tipo de información puede ser aplicada a otra gran variedad de aplicaciones actuales de audio, entre las que destacan las siguientes:

- **Detección de la melodía y bajos:** El objetivo es identificar la línea melódica y la línea de bajos entre un conjunto de posibles pitches activos en la señal polifónica. La línea melódica se caracteriza por tener la estructura armónica predominante en la zona media y alta de frecuencia mientras que la línea de bajos predomina en la zona baja de frecuencia. Su utilidad radica en numerosas aplicaciones como por ejemplo la separación del solista respecto al acompañamiento.
- **Separación de fuentes sonoras:** La separación de fuentes sonoras de audio se puede ver como el siguiente paso a la estimación multi-pitch ya que si un sistema es capaz de identificar los eventos que están activos en cada instante temporal, conocerá la señal producida por cada fuente de audio en cada instante.
- **Manipulación del sonido:** Los efectos de manipulación, edición y remezclado que se pueden aplicar al audio monocanal son totalmente específicos, pudiendo extraer o modificar pistas completas, segmentos variables o eventos individuales al disponer de todos y cada uno de los pitches activos en cada instante temporal.
- **Recuperación de información musical:** serán necesarias aplicaciones de búsqueda, gestión y recuperación de material multimedia. Para satisfacer

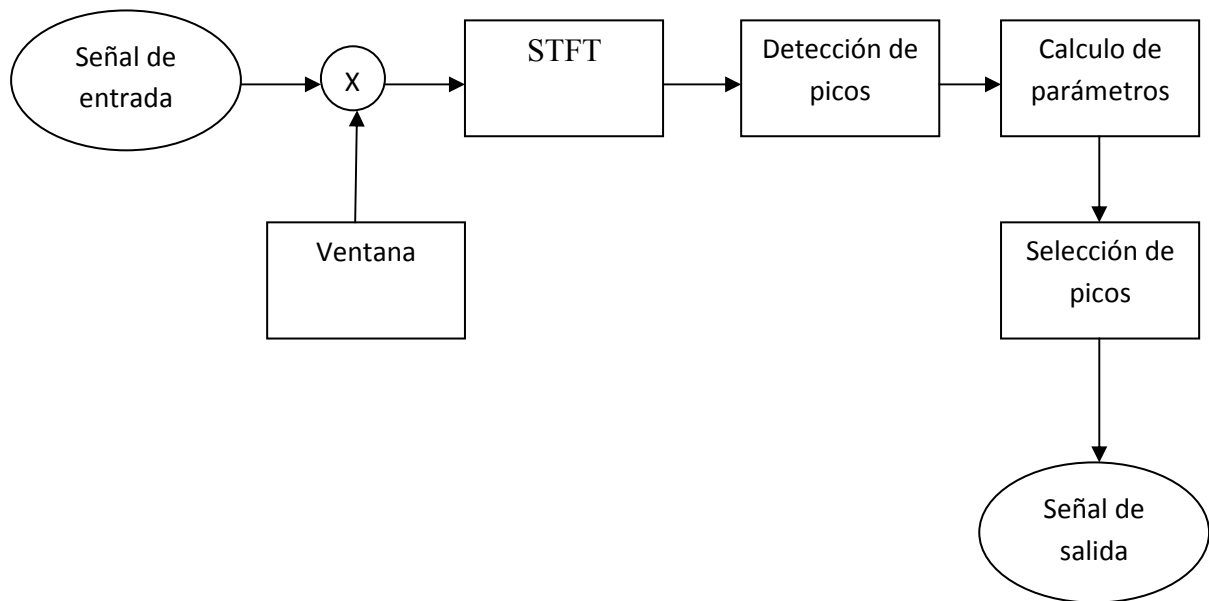
esta necesidad, la información procedente de la secuencia de frecuencias fundamentales extraídas de las señales de audio polifónicas puede ser determinante a la hora de buscar fragmentos de audio con ciertas características.

- Mejora y Cancelación armónica: La mejora armónica se basa en la extracción de una señal periódica (estimación de un pitch determinado) de entre un conjunto de señales ruidosas. En cambio, la cancelación armónica se basa en la extracción de una señal arbitraria de entre un conjunto de señales periódicas. Por tanto, la mejora armónica sólo busca la señal deseada mientras que la cancelación busca las señales consideradas como ruido.

5. ANALISIS Y SINTESIS DE AUDIO

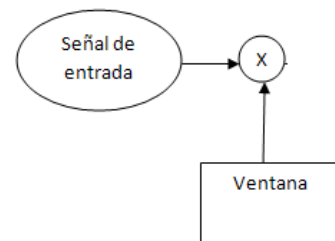
5.1 Esquema General:

El proceso de análisis de audio se resume en el esquema general ilustrado en la figura siguiente:



- División de la señal en tramas y enventanado:

La señal de audio original se divide en tramas de longitud constante a las que se les aplica un enventanado, la ventana seleccionada es la ventana de Hanning.



- STFT:

Mediante la transformada de Fourier se determina el contenido en frecuencia sinusoidal y de fase en cada trama de la señal así como sus cambios con respecto al tiempo.

También permite obtener el espectrograma de la señal.

STFT

- Detección de picos:

Se detectan todos los picos espectrales, ya sean los correspondientes a las sinusoides como los correspondientes al ruido. Para cada trama se almacenan las posiciones de cada uno de los picos.

Detección de
picos

- Calculo de parámetros:

Por cada pico que encuentra la función anterior, calculamos una serie de parámetros, que servirán para crear el descriptor, estos parámetros son: mf (frecuencia central),

Bw (ancho de banda), L (Longitud), B (ancho de banda/Longitud), Bmf ($B \cdot mf$) este ultimo será el descriptor que se usara.

Calculo de
parámetros

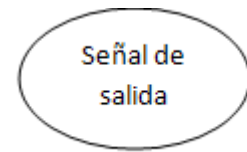
- Selección de picos:

Se usara el descriptor Bmf, para seleccionar solo los picos que pertenezcan al primer armónico de la señal, se representaran los valores del descriptor en un histograma con todos los valores de Bmf.

Selección de
picos

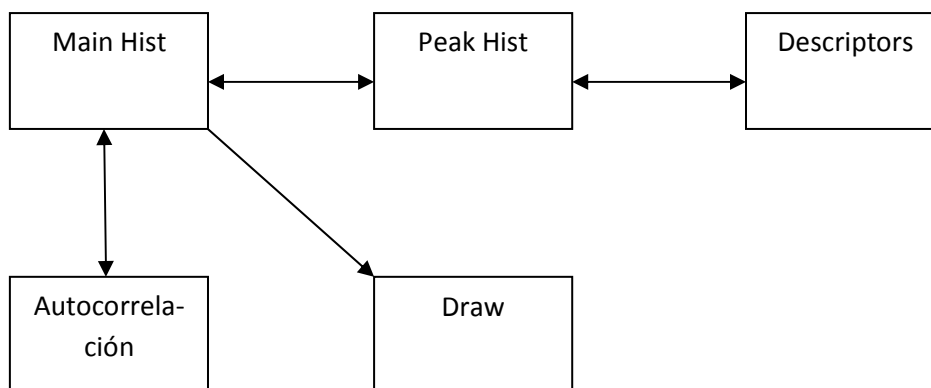
- Señal de salida:

La señal de salida, se corresponderá con el primer armónico de la señal, es decir con la frecuencia fundamental de la señal.



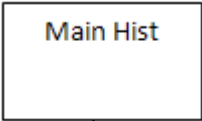
5.2 Esquema particular

Después de ver el esquema general de análisis y síntesis de la señal de audio, a continuación se define el esquema que se utilizara en el programa para calcular la Frecuencia fundamental.



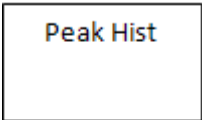
- Main Hist:

Es la función principal del programa, ha ella se le pasa la señal de entrada, así como la frecuencia de muestro y el tamaño y el tipo de ventana a utilizar, es la encargado de llamar a la función peak hist y autocorrelación que calcularan por métodos diferentes la frecuencia fundamental. Una vez que se han calculado las frecuencias fundamentales con ambos métodos llama a la función draw.

A rectangular box with a black border containing the text "Main Hist" in a blue, sans-serif font.

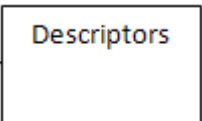
- Peak Hist:

Esta función recibe la señal de entrada junto con la frecuencia fundamental, el tipo de ventana y su tamaño, divide la señal en tramas, que se las va pasando a la función descriptors, y crea el histograma con los valores del descriptor que recibe de descriptors, también es el encargado de crear el espectrograma, le devuelve a la función main hist, el histograma con los valores del descriptor, el espectrograma y la frecuencia fundamental de la señal.

A rectangular box with a black border containing the text "Peak Hist" in a blue, sans-serif font.

- Descriptors:

Esta función es la encargada de calcular todos los parámetros necesarios para crear el descriptor, recibe una trama de la señal y se encarga de detectar todos los picos de esa trama y calcular el valor del descriptor de cada pico, también realiza la transformada de Fourier de cada trama, devuelve a Peak Hist los valores del descriptor y la transformada de Fourier.

A rectangular box with a black border containing the text "Descriptors" in a blue, sans-serif font.

Una vez que toda la señal se ha analizado, recibe los valores del descriptor donde se encuentra la frecuencia fundamental y realiza la misma operación, quedándose únicamente con los picos que se corresponden con la frecuencia fundamental.

- Autocorrelación:

Esta función recibe la señal de entrada y calcula su frecuencia fundamental por el método de autocorrelación y devuelve a la función main hist la frecuencia fundamental.

Autocorrelac
ion

- Draw:

Esta función recibe los valores de la frecuencia fundamental que se han calculado por los dos métodos diferentes y es la encargada de representarlos.

Draw


6. FUNCIONAMIENTO DE LA HERRAMIENTA

6.1 Cálculo de espectrograma:

El primer paso para el análisis consiste en calcular los espectros en magnitud y fase del fragmento de señal centrado en el punto de medida actual. El cálculo del espectro es llevado a cabo por medio de la técnica de la transformada de Fourier en tiempo corto, STFT.

Los parámetros de control para la STFT tales como el tamaño y el tipo de ventana, el número de puntos de la FFT o la frecuencia a la que se toman los puntos de medida, han de ser acordes con el tipo de sonido sometido a estudio. En primer lugar es necesaria una buena resolución frecuencial, pues el proceso ha de ser capaz de identificar los picos espectrales de la señal.

Es importante la correcta elección por parte del usuario del tipo de ventana utilizada para el análisis, ya que determina la resolución en el tiempo y en la frecuencia, afectando directamente a la suavidad del espectro y por tanto condicionando la correcta detección de los picos espectrales. En este proyecto se ha determinado el uso de una ventana tipo Hanning.

Es conocido que al enventanar una serie de sinusoides y aplicar la transformada de Fourier se obtiene como resultado una réplica de la transformada de la ventana centrada en cada una de las frecuencias de las sinusoides. Esto ha de tenerse en cuenta a la hora de estimar correctamente los picos espectrales de la señal, ya que los lóbulos laterales pueden resultar un problema. Interesa por tanto el control de dos parámetros principales de la ventana de análisis: la anchura del lóbulo principal y la relación entre el lóbulo principal y el lóbulo secundario. Las ventanas más utilizadas en las aplicaciones de audio son las de tipo Blackman y Hanning. Se ha optado en el presente proyecto por una ventana de tipo Hanning para todos los análisis ya que presenta un lóbulo lateral estrecho y una relación de potencia entre lóbulos principal y secundario de  30 dB. La ventana tipo Blackman presenta una anchura del lóbulo lateral mayor que la de Hanning además de una mejor relación de potencia entre lóbulos, el problema es que el ruido enmascara los lóbulos laterales.

El tamaño de la ventana es clave también para el buen funcionamiento de la herramienta. La señal a analizar no es estacionaria: contiene componentes sinusoidales de frecuencia, amplitud y fase variables, de modo que un análisis mediante el uso de STFT requiere la elección de una ventana lo suficientemente corta como para que el fragmento de la señal que encierra pueda considerarse pseudo-estacionario. Al mismo tiempo, una ventana de análisis corta tiene un gran ancho de banda en el dominio frecuencial, lo que desemboca en una pérdida de discriminación de picos por el solapamiento de las replicas de la ventana en los picos.

6.2 Clasificación de picos espectrales:

Tras haber realizado la transformada corta de Fourier, STFT, sobre una trama se procede a la detección de picos. La herramienta devuelve una lista con las posiciones de cada pico. Tras tener localizados todos los picos y tener bien definida la información espectral de estos se procede al cálculo de los parámetros correspondientes a cada pico, el programa calcula 5 parámetros:

- Frecuencia media (mf): Por cada pico que encuentra calcula la frecuencia central de dicho pico.
- Ancho de banda (Bw): Este parámetro proporciona el ancho de banda del pico seleccionado.
- Longitud (L): Este parámetro devuelve la longitud que tiene el pico, es decir el número de puntos que lo forman.
- Ancho de banda normalizado (B): Este parámetro es la división del Ancho de banda (Bw) entre la longitud (L).
- Estos parámetros ya se estudiaron con anterioridad en otro proyecto, aparece en la bibliografía con la referencia [3].

Estos parámetros se basan en la teoría del ancho de banda normalizado que dice lo siguiente.

Para un cierto pico de un armónico en el dominio frecuencial se tiene un cierto ancho de banda (Bw) y una cierta longitud de pico (L), pues bien si a esta misma señal se le aplica una modulación en amplitud o en frecuencia su espectro cambiara de modo que el pico será más ancho con lo que se obtendrá un ancho de banda (Bw) mayor y una longitud de pico (L) mayor, pero el cociente entre ellos (Bw/L) será no constante pero si muy parecido.

Con lo que este parámetro Ancho de banda normalizado sirve para diferenciar los picos de los armónicos, de los picos de ruido, ya que los picos de ruido son de una forma arbitraria, cada uno diferente con un ancho de banda (Bw) y una longitud (L) diferente para cada pico y por tanto un cociente diferente en cada caso.

En la siguiente figura se puede ver más claramente.

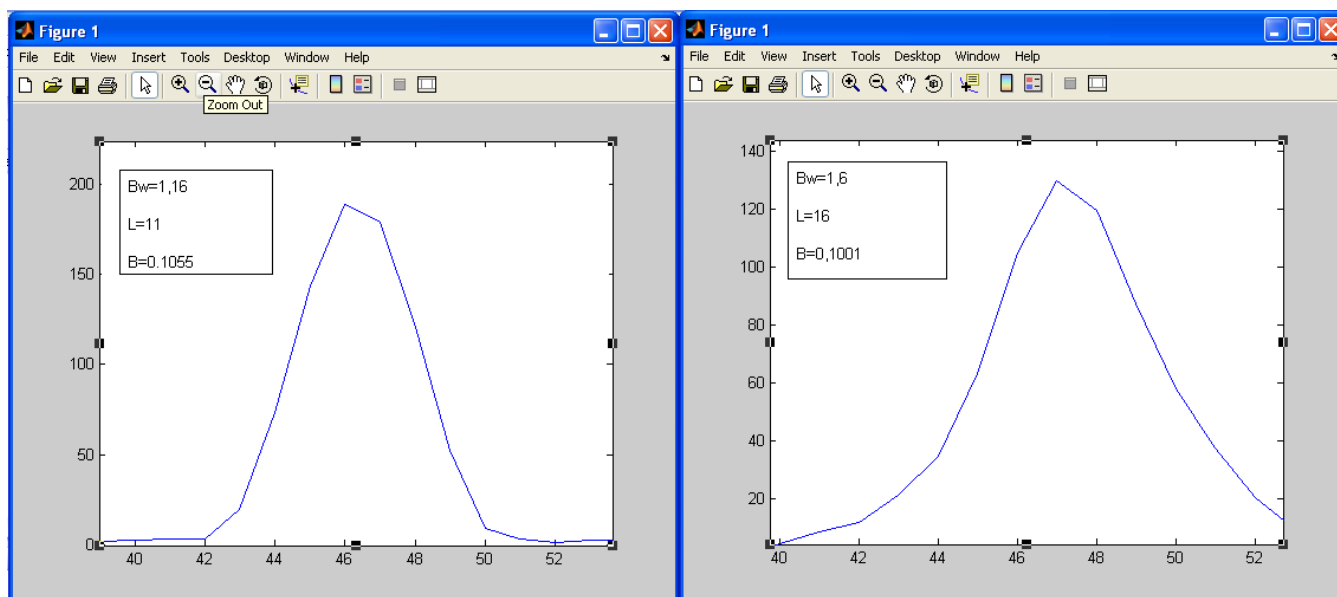


Figura 6.1: La figura de la izquierda muestra el pico de una señal compuesta por un tono puro. La figura de la derecha es la misma señal que la de la izquierda aplicándole una modulación en amplitud y en frecuencia.

Como se observa en la figura 6.1 al modular una señal en frecuencia y amplitud su espectro se ensancha, con lo que los parámetros Bw y L aumentan, en el primer caso tenemos una Bw de 1,16 y una L de 11, pero al modular la señal estos parámetros aumentan ya que se obtiene una $Bw=1,6$ y una L de 16.

En cambio si se compara el valor de B (Bw/L), en los dos casos es muy parecido 0,1055 de la primera señal, respecto a 0,1001 de la segunda, con lo este parámetro nos permitirá discernir entre picos espectrales de armónicos y picos de ruido.

- Descriptor (Bmf): Es simplemente la multiplicación del ancho de banda normalizado por la frecuencia media del pico.

Al multiplicar el valor del ancho de banda normalizado (Bw/L) por la frecuencia central del pico (mf) se consigue agrupar los valores de los diferentes armónicos, si previamente se había conseguido separar el valor de los picos armónicos de los picos de ruido, con este parámetro cada armónico fundamental tendrá un rango de valores característico, que se sabrá al mostrar un histograma con todos los valores de del descriptor Bmf

A partir de este histograma se sacaran los valores del descriptor donde se encuentran los picos correspondientes a la frecuencia fundamental, que serán los valores del primer pico.

7. RESULTADOS EXPERIMENTALES

7.1 Procedimiento

Se pretende realizar un estudio comparativo entre una técnica nueva basada en un descriptor a través del cual se calculara la frecuencia fundamental y una técnica ya existente conocida como autocorrelación que calcula la frecuencia fundamental de una señal a través de una serie de parámetros.

Los diferentes casos a estudiar son:

- Calculo de frecuencia fundamental de una señal simple (Una sola fuente).
- Calculo de la frecuencia fundamental de una señal múltiple (Dos fuentes).

El método de actuación para la comparación de las diferentes técnicas será el siguiente: Se seleccionara una señal de audio, se representará su espectrograma para ver la composición de sus diferentes armónicos, el primero de ellos se corresponderá con la frecuencia fundamental.

A través del descriptor Bmf se seleccionaran solo con los picos de la señal que pertenezcan al primer armónico, obteniendo de este modo la frecuencia fundamental.

Por otra parte con el método de autocorrelación se obtendrá directamente el pitch de la señal, pudiendo compararlos de este modo.

7.2 Calculo de la frecuencia fundamental a través del descriptor:

Para el cálculo de la frecuencia fundamental en este primer caso se utilizara la siguiente señal, es la nota de una guitarra eléctrica, la señal esta muestreada a 44100 muestras por segundo y se encuentra codificada con 16 bits por muestra, consta de un canal, tiene la siguiente forma:

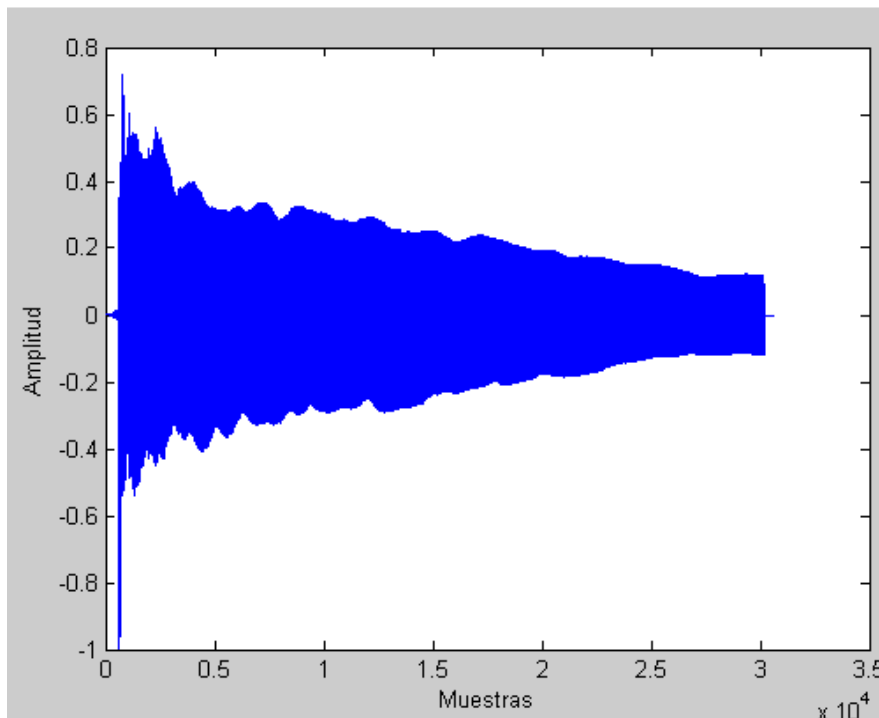


Figura 7.1: Señal compuesta por una nota de guitarra que utilizaremos para el cálculo de la frecuencia fundamental.

Al mostrar su espectrograma se puede observar la distribución de sus armónicos a través del eje frecuencial, así como su frecuencia fundamental, la amplitud de los diferentes armónicos y la separación entre ellos así como su limite frecuencial:

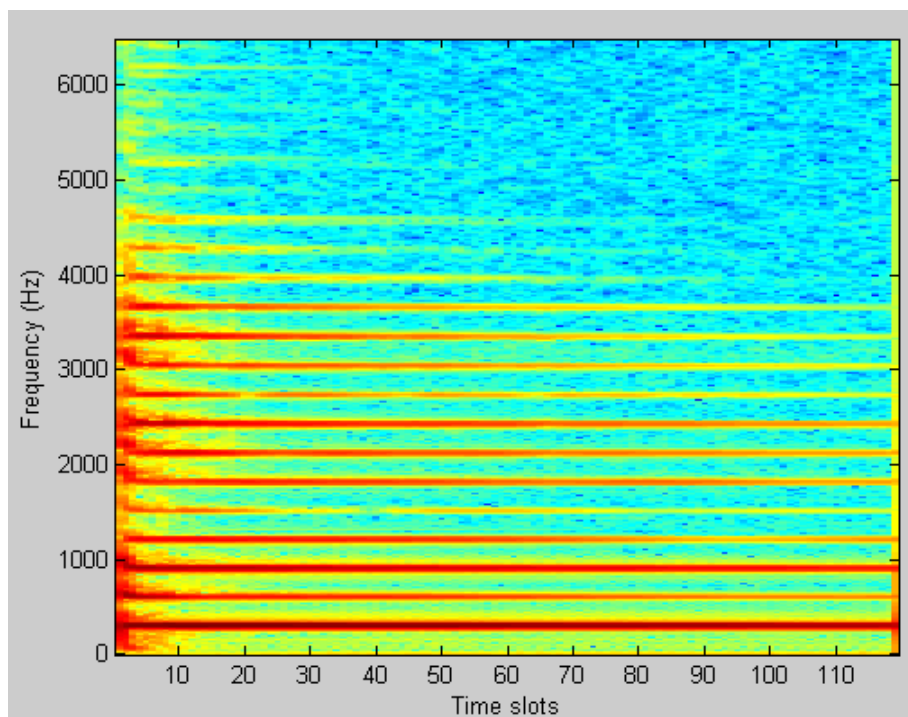


Figura 7.2: Espectrograma de la señal compuesta por una nota de guitarra que utilizaremos para el cálculo de la frecuencia fundamental.

La frecuencia fundamental se corresponderá con el primero de los armónicos, que en este caso está en torno a los 300 Hz, se puede observar que se diferencian hasta 11 armónicos que llegan hasta la frecuencia de 4000 Hz, la amplitud de los diferentes armónicos va decreciendo conforme aumenta la frecuencia.

También se observa que aparece algo de ruido al principio de la señal, lo que dificultara la separación de la frecuencia fundamental.

Para el cálculo de la frecuencia fundamental se ha construido un descriptor al que denominaremos Bmf, que se basara en el criterio del ancho de banda normalizado, este descriptor calculara en cada pico que tenga la señal la siguiente operación:

$$B = \frac{BW}{L} < 1$$

Bw: Ancho de banda del pico seleccionado

L: Longitud del pico seleccionado

$$Bmf = B \cdot mf$$

mf: Frecuencia central del pico seleccionado

El valor de B estará comprendido entre 0 y 1, será pequeño para picos armónicos y elevado para los picos no armónicos y al multiplicar este valor por la frecuencia central del pico obtendremos unos valores muy similares para todos los picos que pertenezcan al primer armónico, que son los que necesitamos para hallar la frecuencia fundamental.

A continuación realizaremos un histograma con todos los valores de este descriptor obteniendo la siguiente figura:

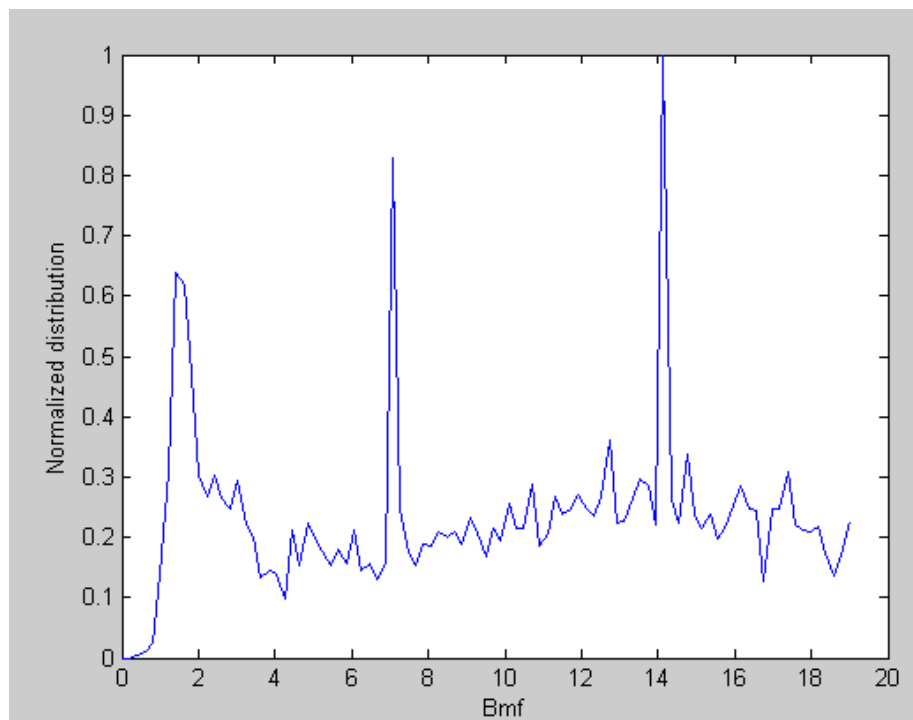


Figura 7.3: *Histograma con los valores del descriptor Bmf, los valores están normalizados de 0 a 1.*

En este histograma se diferencian claramente tres picos diferentes suficientemente separados que se corresponderían con los valores del descriptor de los tres primeros armónicos, como el único que interesa es el primero, se seleccionaran únicamente los picos de la señal cuyo descriptor Bmf tenga unos valores que entren dentro del primer pico de este histograma.

En este caso los valores correspondientes al descriptor Bmf que pertenecerían al primer armónico irían aproximadamente de 1 a 1,8, esos son por tanto los valores que interesan para conseguir la frecuencia fundamental.

La señal se va analizando trama a trama, y en cada trama realiza la siguiente operación:

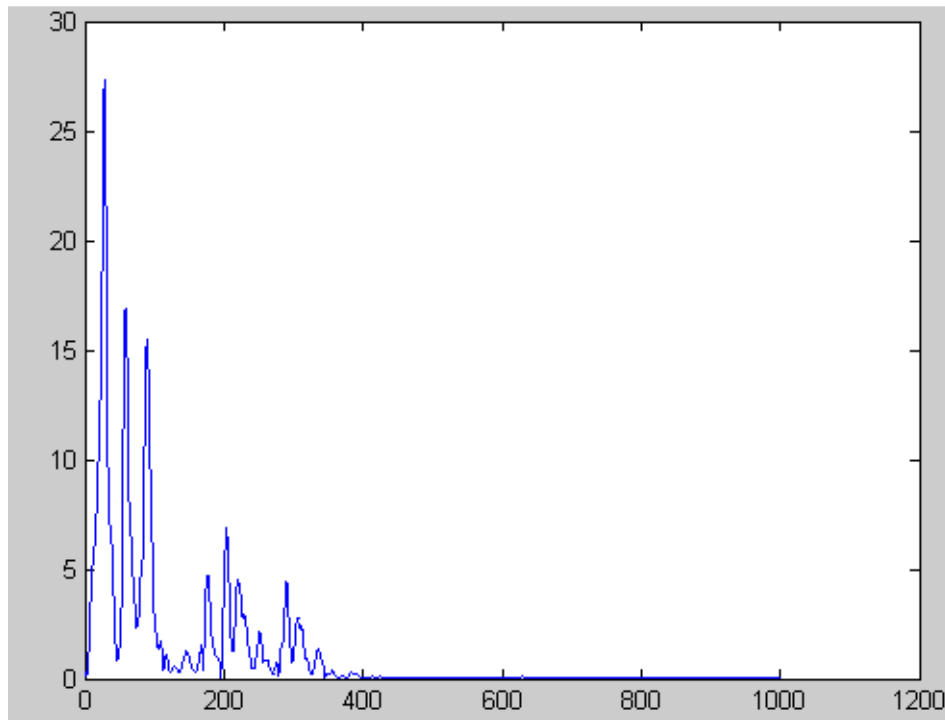


Figura 7.4: *Transformada de Fourier de la primera trama de la señal.*

Esta señal se corresponde con la transformada de Fourier de la primera trama de la señal, podemos observar los diferentes picos de la señal que se corresponderán con los armónicos de la señal, los 3 primeros armónicos se diferencian perfectamente, pues bien el programa una vez obtenidos los valores del descriptor, anteriormente calculados, que se corresponden con el primer armónico realiza la siguiente operación:

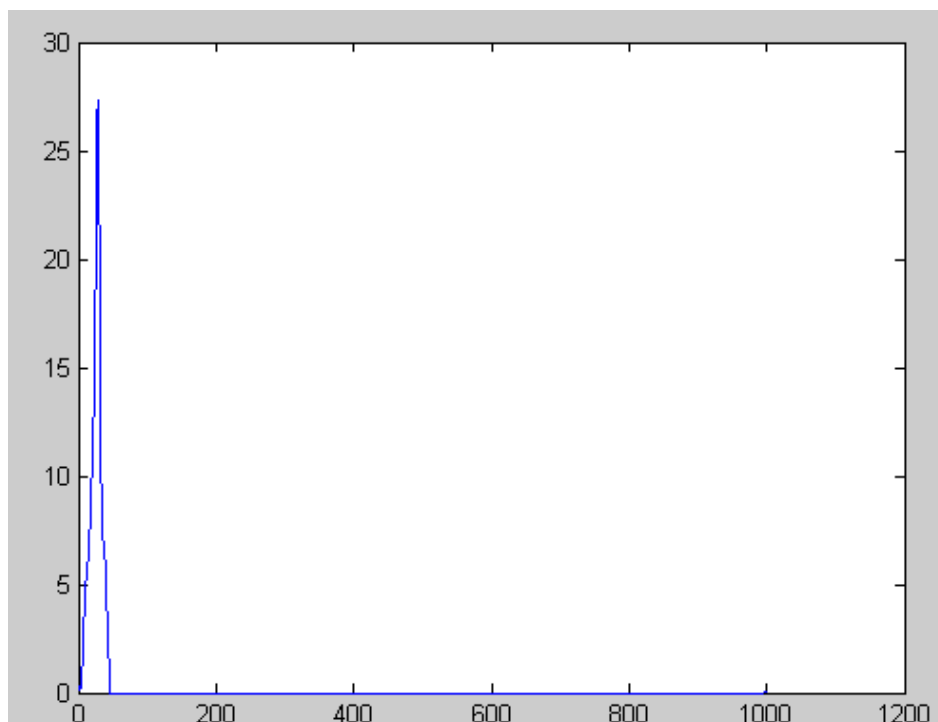


Figura 7.5: *Transformada de Fourier de la primera trama de la señal una vez hemos utilizado el método del descriptor.*

Se observa que el programa a eliminado todos los picos menos el primero que es el que verdaderamente interesa, esto se debe a que los valores del descriptor que no estén dentro del rango que anteriormente se ha calculado no son procesados, con lo que una vez se hayan analizado todas las tramas, se obtendrán únicamente los valores del primer armónico de la señal es decir el valor de la frecuencia fundamental de la señal, y únicamente se deberá representar su espectrograma en función del tiempo y la frecuencia.

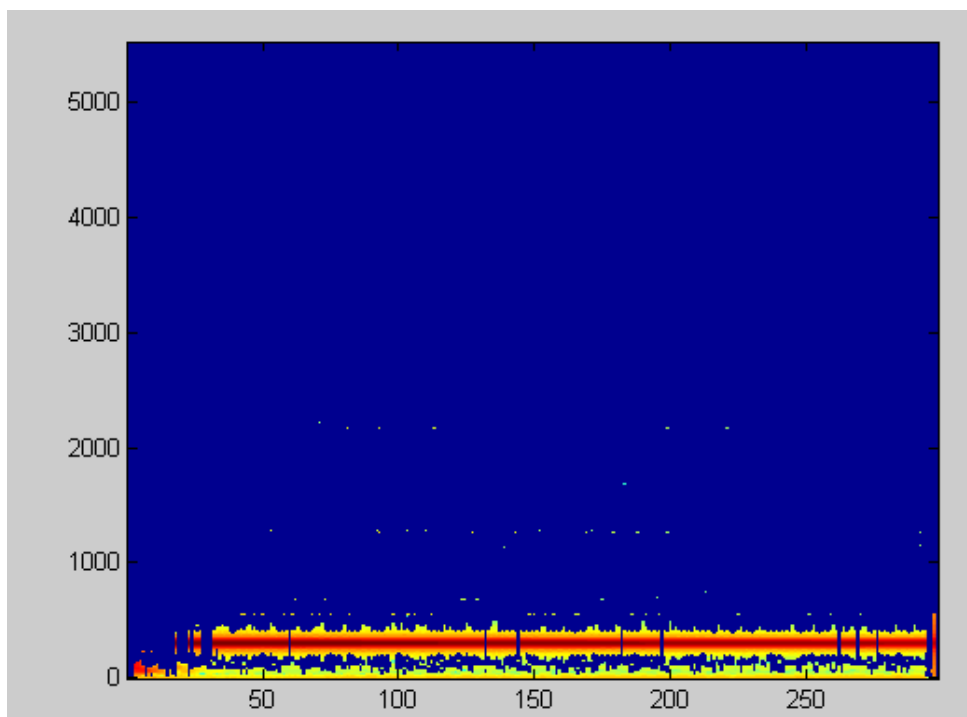


Figura 7.6: *Espectrograma de la señal una vez utilizado el criterio del descriptor.*

En esta imagen se observa como el programa ha eliminado todos los picos de la señal que no cumplían con el criterio establecido del descriptor Bmf, y se ha quedado únicamente con los valores correspondientes al primer armónico, obteniendo por tanto la frecuencia fundamental que tiene un valor aproximado de 300 Hz.

La primera parte de la señal no ha quedado representada a causa del ruido, ya que el programa no es capaz de separar esos valores de los de la frecuencia fundamental del ruido y también porque la señal no empieza en el instante cero. También se puede observar que aparece ruido en la parte inferior debido a la componente DC.

Se podría mejorar esta representación, eliminando los puntos aislados y el ruido que aparece en el espectrograma con un post-procesado de imagen este método consistiría en lo siguiente.

El primer paso es elegir un tamaño de celda por ejemplo 4x5 bins e ir desplazándola por todo el espectrograma, en cada punto se calculara la energía de dicha celda, si la energía es pequeña, se convertirán esos bins en cero, porque se tratara de un punto aislado o de ruido que no pertenece a la frecuencia fundamental, por el contrario si la energía es alta se dejaran esos bins con el mismo valor, de modo que se obtendrá una señal más limpia.

Para poder comparar ambos métodos, primero hay que pasar de la figura del espectrograma 7.6 a una figura parecida a la de autocorrelación, por lo que se mostraran todas las frecuencias centrales de cada uno de los picos considerados del primer armónico.

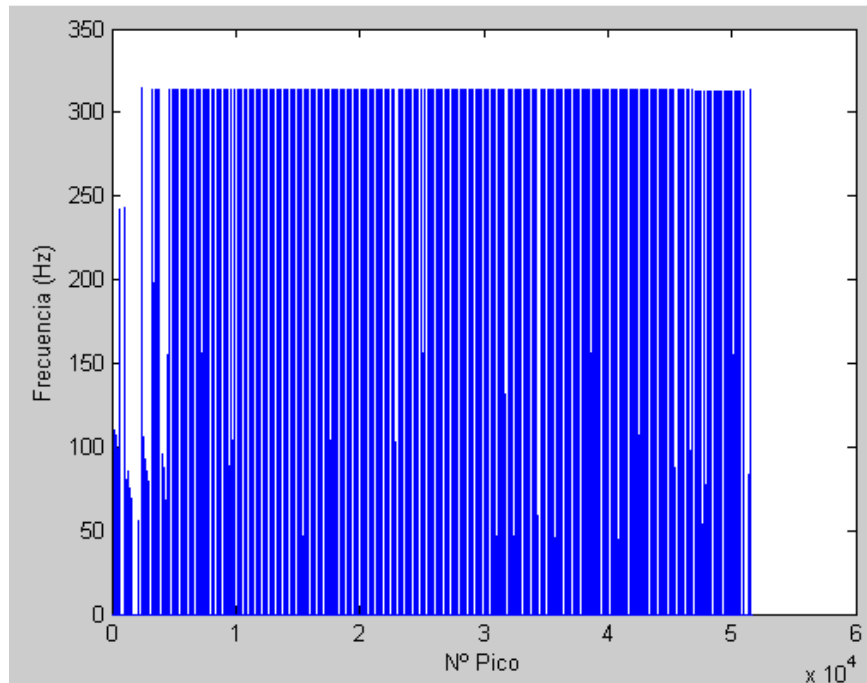


Figura 7.7: *Pitch de la señal de guitarra utilizando el método del descriptor, en concreto es la frecuencia central de cada uno de los picos que se consideran que pertenecen al primer armónico.*

Se obtiene un pitch aproximado de unos 310 Hz un valor idéntico al que se obtiene en el espectrograma de la figura 7.6.

Se puede observar que aparecen algunas frecuencias bajas, esto es debido a los picos de ruido que aparecen en la parte inferior debido a la componente DC y al ruido introducido por el instrumento.

Estas frecuencias que no corresponden a la frecuencia fundamental se podrían eliminar de una manera relativamente sencilla, simplemente habría que comparar cada una de ellas con la frecuencia fundamental en este caso 310 y si la varianza entre ellas es suficientemente alta se podría eliminar, a continuación únicamente habría que interpolar linealmente los valores restantes.

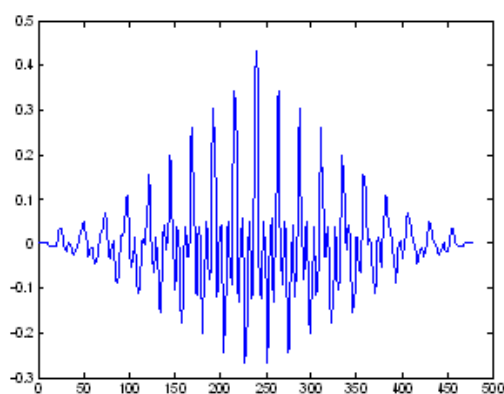
7.3 Cálculo de la frecuencia fundamental a través del método de autocorrelación:

El método con el que se van a comparar los resultados obtenidos, es con el método de autocorrelación:

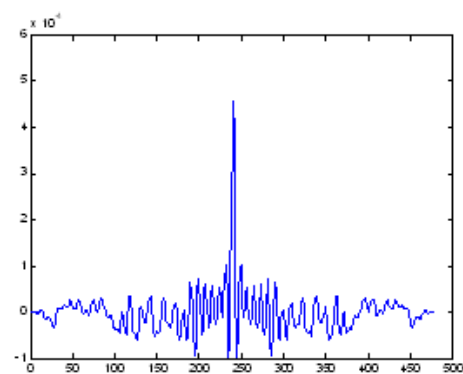
Para ello se divide la señal en tramas con lo que se va enventanando la señal, a continuación se elimina el valor medio de cada trama, se realiza un filtro paso bajo y se calcula la autocorrelación de esta señal.

Una vez se tiene la señal de autocorrelación se puede distinguir entre dos casos, analizando el pico central que siempre será el pico de mayor amplitud y el segundo pico máximo. Una vez se han localizado los dos picos guardaremos su amplitud y posición.

- Trama sonora: La trama se considerara sonora si la amplitud del segundo pico máximo es mayor que el 25% del primer máximo, en este caso se procederá a calcular el pitch, que será el valor de la posición del segundo pico máximo mas el valor de la amplitud de dicho pico.
- Trama sorda: La trama se considerara sorda si la amplitud del segundo pico máximo es menor que el 25% del primer máximo, en este caso el pitch será cero.



Autocorrelación trama sonora.



Autocorrelación trama sorda.

Figura 7.8: La primera imagen representa un ejemplo de la técnica de autocorrelación de una trama sonora, mientras que en la segunda imagen es una trama sorda

En la imagen 7.8 es un ejemplo de autocorrelación de una trama sonora y de otra sorda, se pueden diferenciar fijándose en el segundo pico máximo, en la primera señal este pico está muy cerca del pico máximo central y su amplitud es elevada, mientras en la segunda imagen se aprecia como el segundo máximo tiene un valor mucho más pequeño que el pico máximo central. Por lo que se considerara esta trama como sorda.

Volviendo a la señal de guitarra si se calcula la frecuencia fundamental con el método de autocorrelación se obtienen los siguientes resultados:

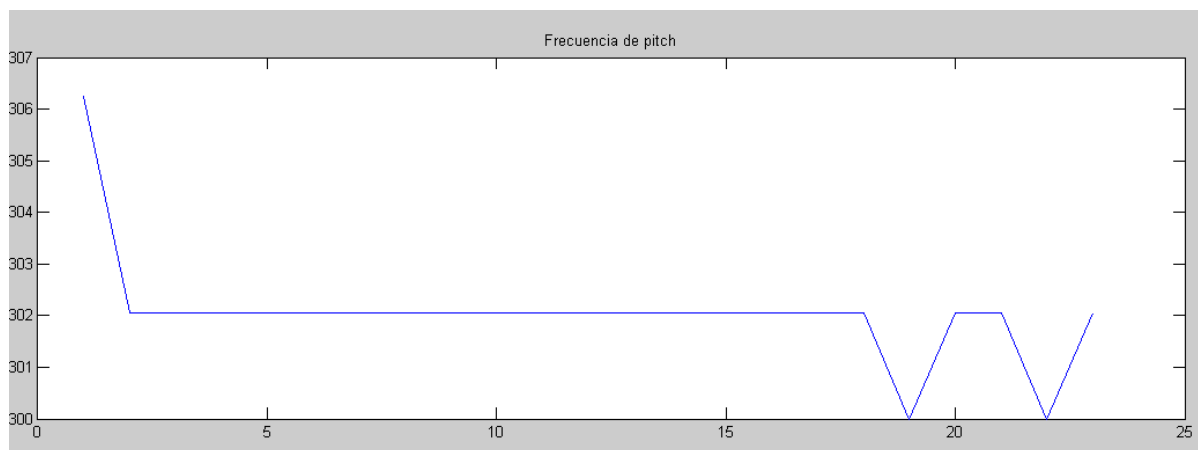


Figura 7.9: *Pitch de la señal de guitarra utilizando el método de autocorrelación.*

Se puede observar que el método de autocorrelación nos da un pitch aproximado de 300 Hz.

Este método da un valor por cada trama, se puede ver que da prácticamente un valor constante.

7.4 Comparación de Métodos para señal con una sola fuente:

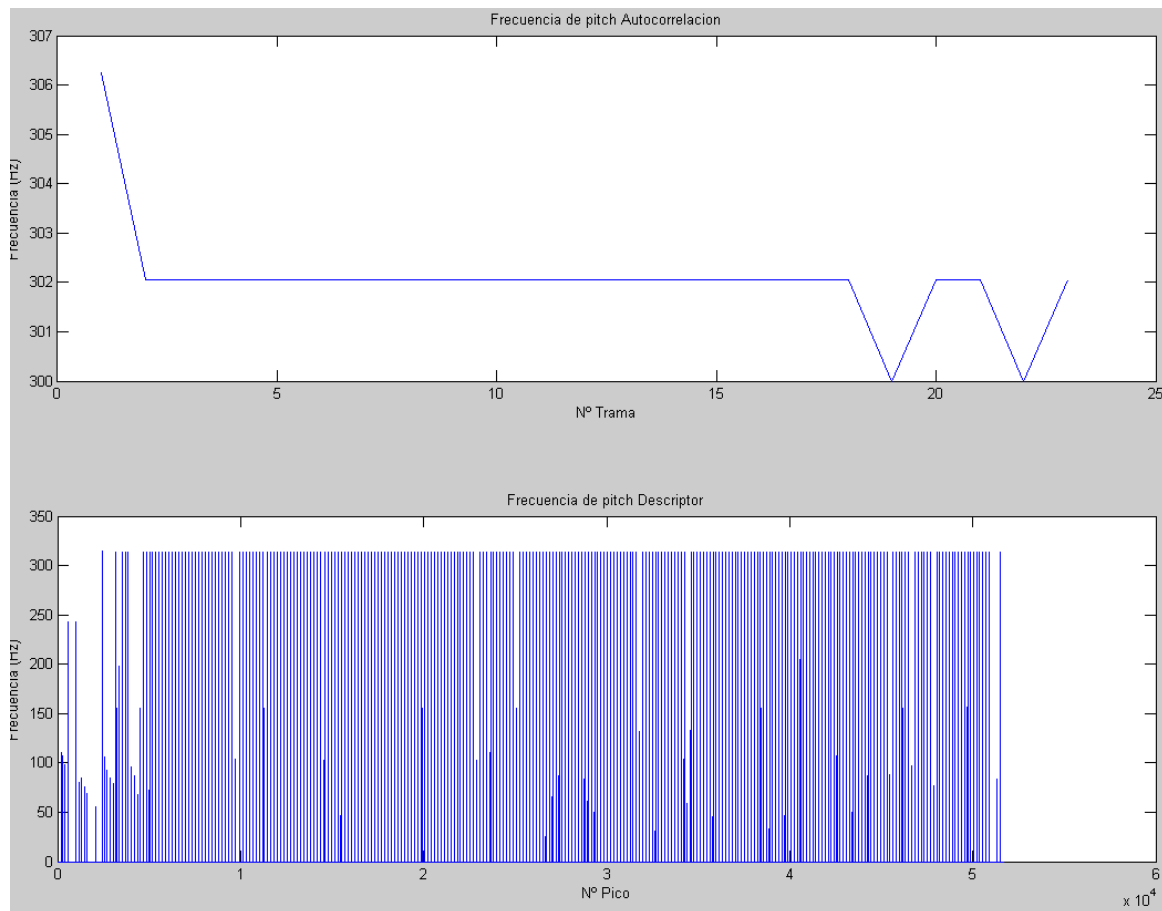


Figura 7.10: La figura superior es la frecuencia fundamental que se obtiene por medio de la autocorrelación. La figura inferior es la frecuencia fundamental que se obtiene con el método del descriptor.

A primera vista se observan algunas diferencias, el método de autocorrelación da un valor para cada trama, mientras que con el método del descriptor se tiene un valor por cada pico que pertenece al armónico fundamental.

Sin embargo se puede concluir que los dos métodos funcionan bien para una señal con una sola fuente, ya que sus valores son muy similares con el método de autocorrelación obtenemos un pitch de 302 Hz, mientras que con el método del descriptor obtenemos un pitch unos 310 Hz.

A continuación se pasara a analizar una señal múltiple con dos fuentes y se probaran los mismos métodos para ver si ambos son eficaces.

7.5 Calculo de la frecuencia fundamental a través del método del descriptor para una señal con dos fuentes:

Para este caso se utilizan dos señales con una única fuente cada una y se sumaran obteniendo una señal múltiple. Se utiliza una señal de violín y otra de un Oboe, ambas señales están muestreadas a 44100 muestras por segundo y se encuentran codificadas con 16 bits por muestra, constan de un canal, tienen la siguiente forma:

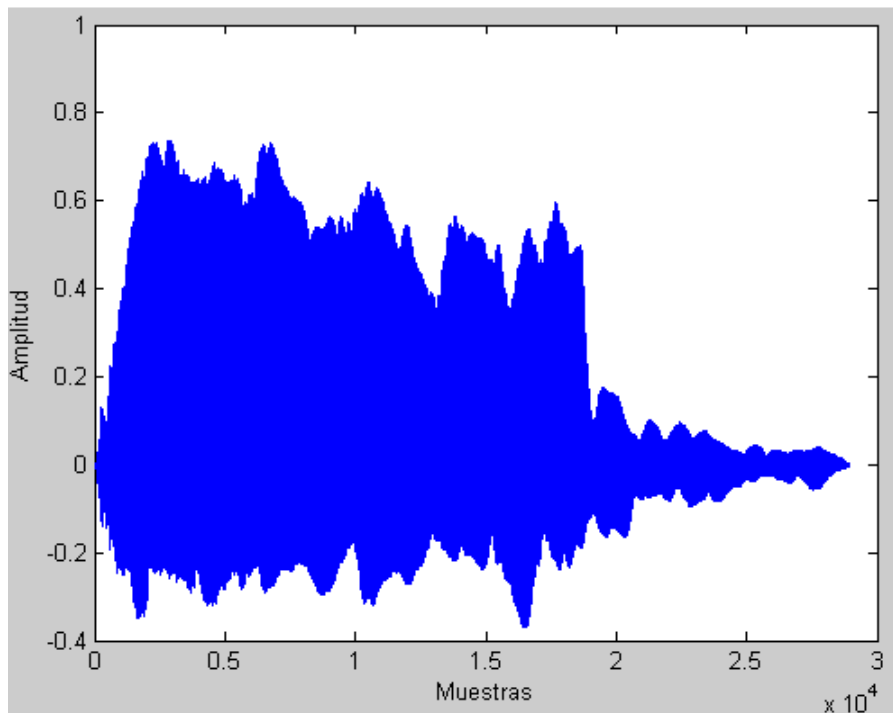


Figura 7.11: Señal compuesta por una nota de violín a la que se sumara la siguiente señal que se puede ver en la figura 7.12

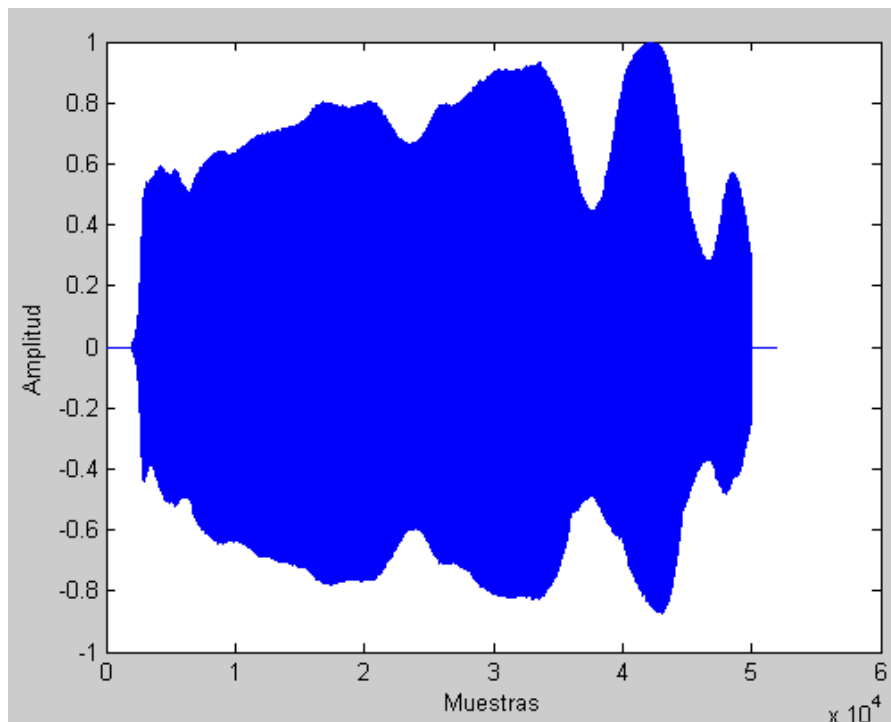


Figura 7.12: Señal compuesta por una nota de Oboe a la que se sumara la señal de la figura 7.11.

Si se muestran sus espectrogramas se puede observar la separación de sus diferentes armónicos y así poder analizar mejor la señal:

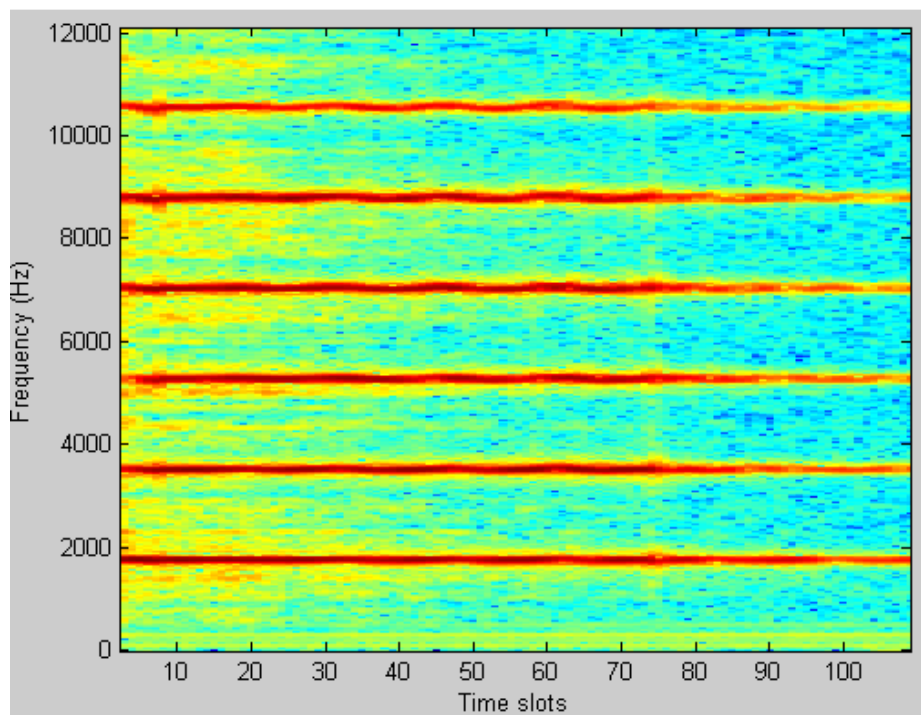


Figura 7.13: Espectrograma de la señal de violín (figura 7.11)

La frecuencia fundamental se corresponderá con el primero de los armónicos, que en este caso está en torno a los 1900 Hz, se puede observar que se diferencian hasta 6 armónicos que llegan hasta la frecuencia de 11000 Hz. La frecuencia fundamental para esta señal es alta ya que se trata de un violín cuyo rango de frecuencias puede ir de 200 a 3500 Hz, los armónicos se encuentran situados a frecuencias múltiplos de la frecuencia fundamental.

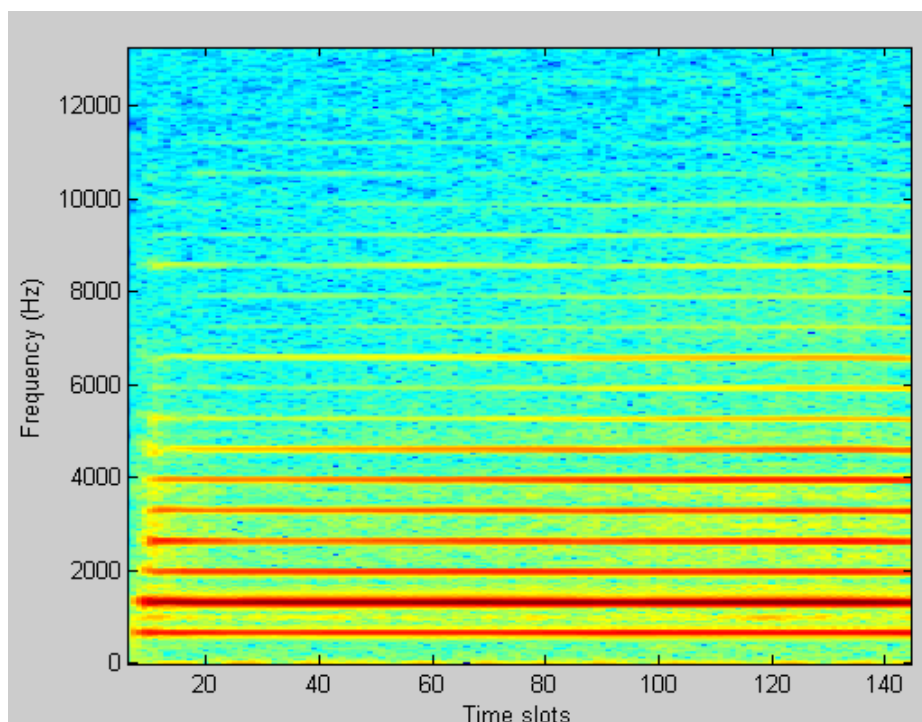


Figura 7.14: *Espectrograma de la señal de Oboe (figura 7.12)*

La frecuencia fundamental se corresponderá con el primero de los armónicos, que en este caso está en torno a los 500 Hz, se puede observar que se diferencian hasta 10 armónicos que llegan hasta la frecuencia de 8000 Hz, la amplitud de los diferentes armónicos va decreciendo conforme aumenta la frecuencia. La frecuencia fundamental de esta señal es más baja ya que el instrumento es un Oboe cuyo rango de frecuencias es menor que el del violín ya que va de 250 a 1500 Hz. Los armónicos se encuentran situados a frecuencias múltiplos de la frecuencia fundamental.

A continuación se muestra la suma de ambos espectrogramas:

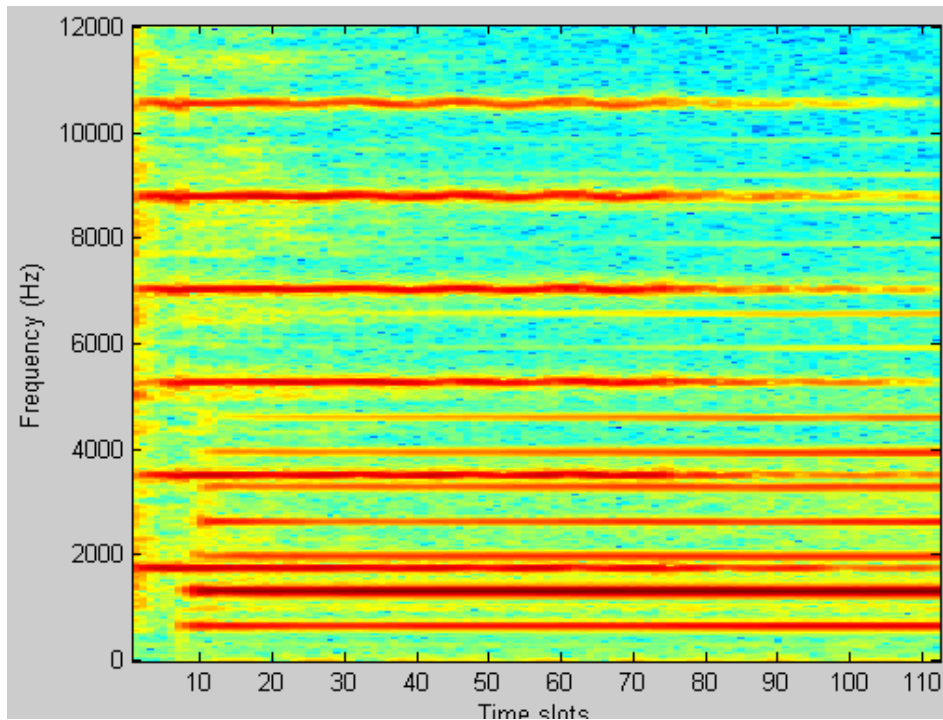


Figura 7.15: *Espectrograma de la señal de violín mas señal del oboe (suma figuras 7.13 más 7.14)*

Se pueden diferenciar ambas señales ya que los armónicos del Oboe empiezan algo más tarde que los del violín y los armónicos de una y otra señal están colocados a frecuencias múltiplos de la frecuencia fundamental. Los armónicos del violín están colocados a frecuencias más altas mientras que los de los Oboe se concentran a frecuencias más bajas.

En el caso de señales múltiples se va a intentar obtener la frecuencia fundamental más baja de los dos instrumentos, que en este caso sería el del oboe que está en torno a los 600 Hz.

A continuación se calcularán los valores del descriptor Bmf para esta señal múltiple y se representará su histograma.

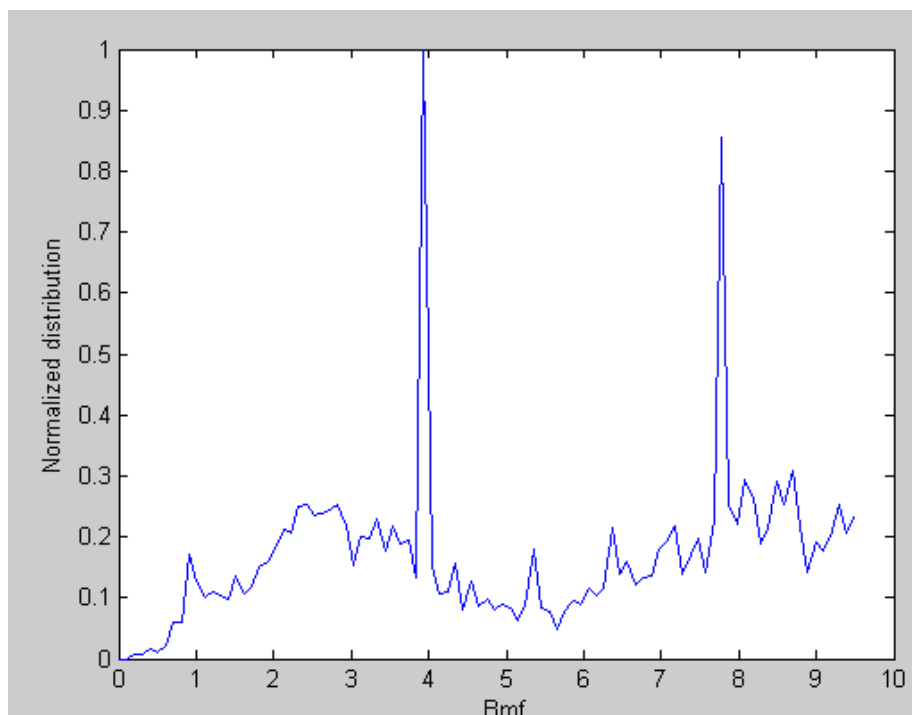


Figura 7.16: *Histograma con los valores del descriptor Bmf, los valores están normalizados de 0 a 1.*

Se pueden observar 2 picos claramente diferenciados, pero solo interesa el primero que será el que correspondiente a los valores del descriptor que pertenezcan al primer armónico de la señal con frecuencia fundamental más baja, en este caso la del oboe.

En este caso los valores correspondientes al descriptor Bmf que pertenecerían al primer armónico irían de aproximadamente de 3,7 a 4,1, esos son por tanto los valores que interesan para conseguir la frecuencia fundamental.

Para comprobar si se está seleccionando el armónico correcto vamos a representar de nuevo el espectrograma de la señal múltiple exceptuando los valores del descriptor que caigan dentro de este primer pico.

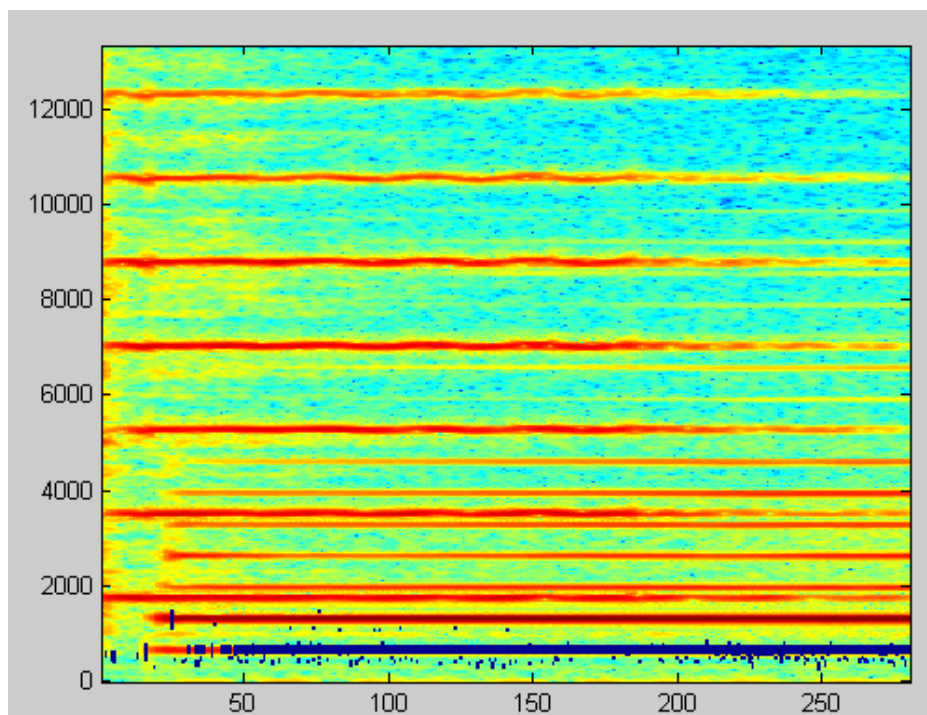


Figura 7.17: *Espectrograma de la señal de violín mas señal del oboe, exceptuando los valores cuyo descriptor Bmf caen dentro del primer pico del histograma de la figura 7.16.*

Se puede observar como los valores del descriptor que están dentro del primer pico del histograma son los que pertenecen al primer armónico de la señal con frecuencia fundamental más baja, en este caso la del oboe, si se representan estos valores se obtendrá la frecuencia fundamental de dicha señal.

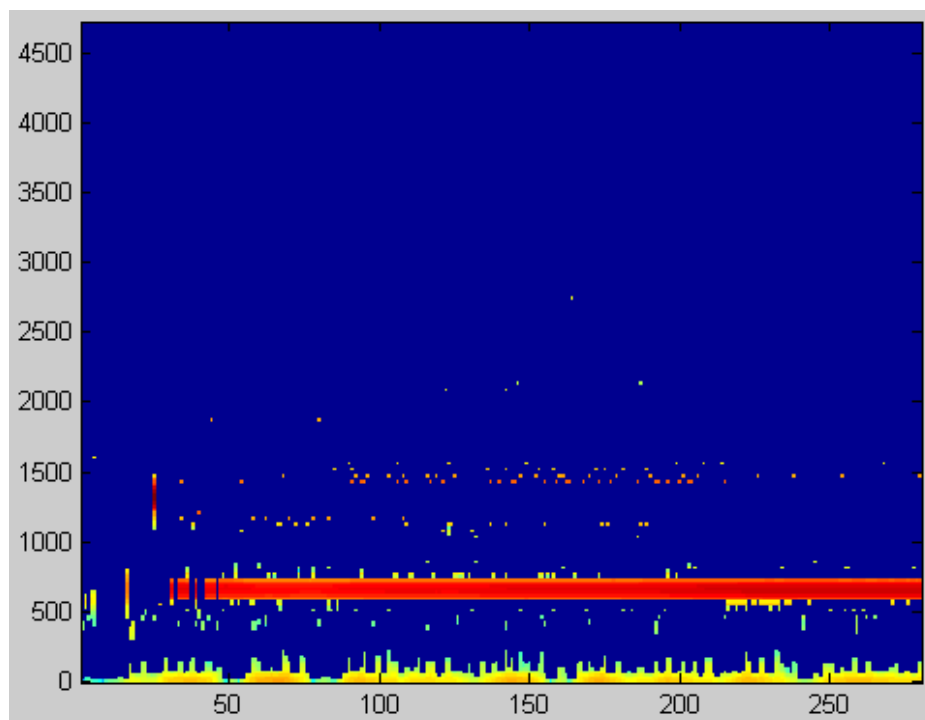


Figura 7.18: *Espectrograma de la señal múltiple una vez utilizado el criterio del descriptor.*

Se puede observar que la frecuencia fundamental se sitúa en torno a los 600 Hz, y que dicha frecuencia corresponde a la señal del Oboe, que es la que tiene el pitch más bajo de la dos señales.

La primera parte de la señal no ha quedado representada a causa del ruido, ya que el programa no es capaz de separar esos valores de los de la frecuencia fundamental. También se puede observar que aparece ruido en la parte inferior debido a la componente DC.

Como se ha comentado anteriormente podemos limpiar esta imagen con un post-procesado de imagen, de modo que se eliminen los puntos aislados y parte del ruido que aparece.

Para ver la frecuencia fundamental de manera más clara se mostrara la frecuencia central de cada pico perteneciente al primer armónico.

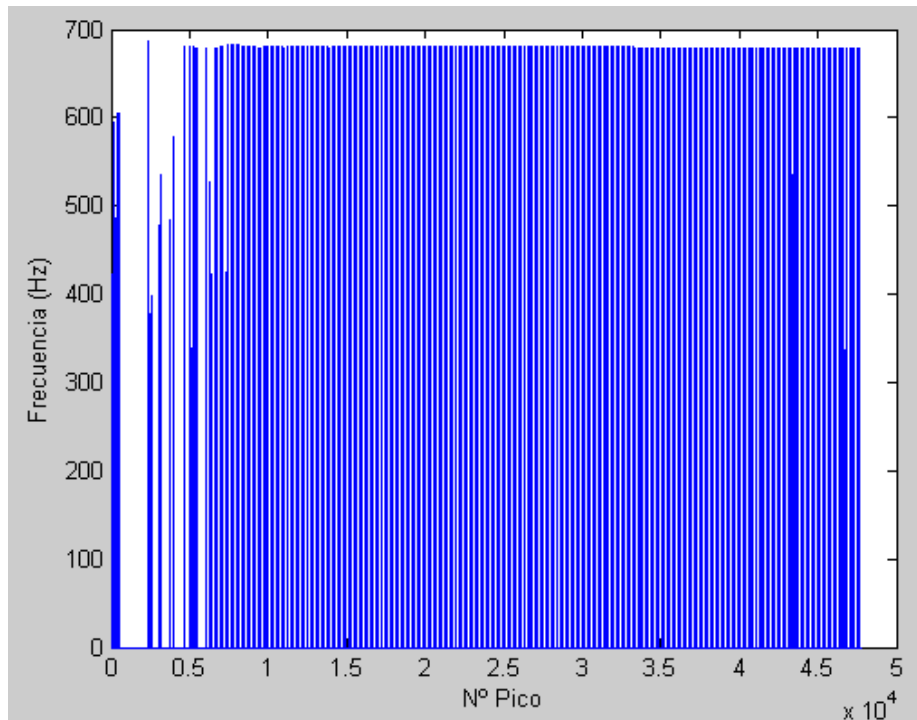


Figura 7.19: *Pitch de la señal utilizando el método del descriptor, frecuencia central de cada uno de los picos que se considera que pertenecen al primer armónico.*

Se puede observar que aparecen frecuencias bajas, esto es debido a los picos de ruido que aparecen en la parte inferior debido a la componente DC y al ruido introducido por los instrumentos.

Obtenemos un pitch unos 680 Hz un valor idéntico al que se obtiene en el espectrograma de la figura 7.18.

7.6 Calculo de la frecuencia fundamental a través del método de autocorrelación para una señal con dos fuentes:

Se realizar la misma operación de autocorrelación que hemos utilizado para la señal simple para comprobar si se obtienen unos resultados similares.

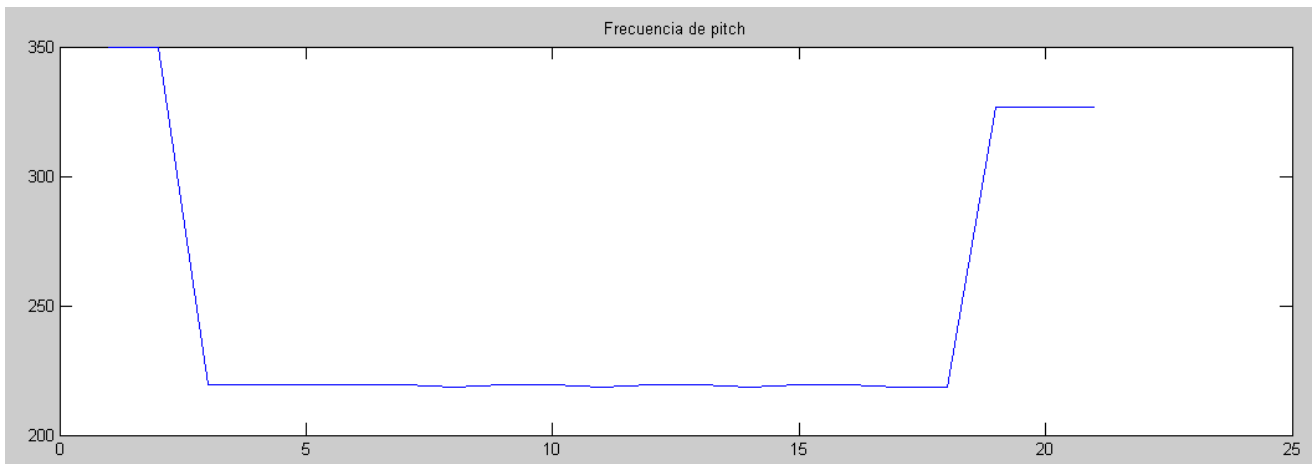


Figura 7.20: *Pitch de la señal múltiple utilizando el método de autocorrelación.*

Para la señal multiple se obtiene un pitch aproximado de 220 Hz, aunque hay tramas que nos dan un valor superior, la primera parte de la señal tiene un valor de 350 Hz, mientras que la ultima parte tiene un valor de 320 Hz.

Estos valores distan muxo de los que deberian salir mirando el espectrograma de la señal.

7.7 Comparación de Métodos para señal con dos fuentes:

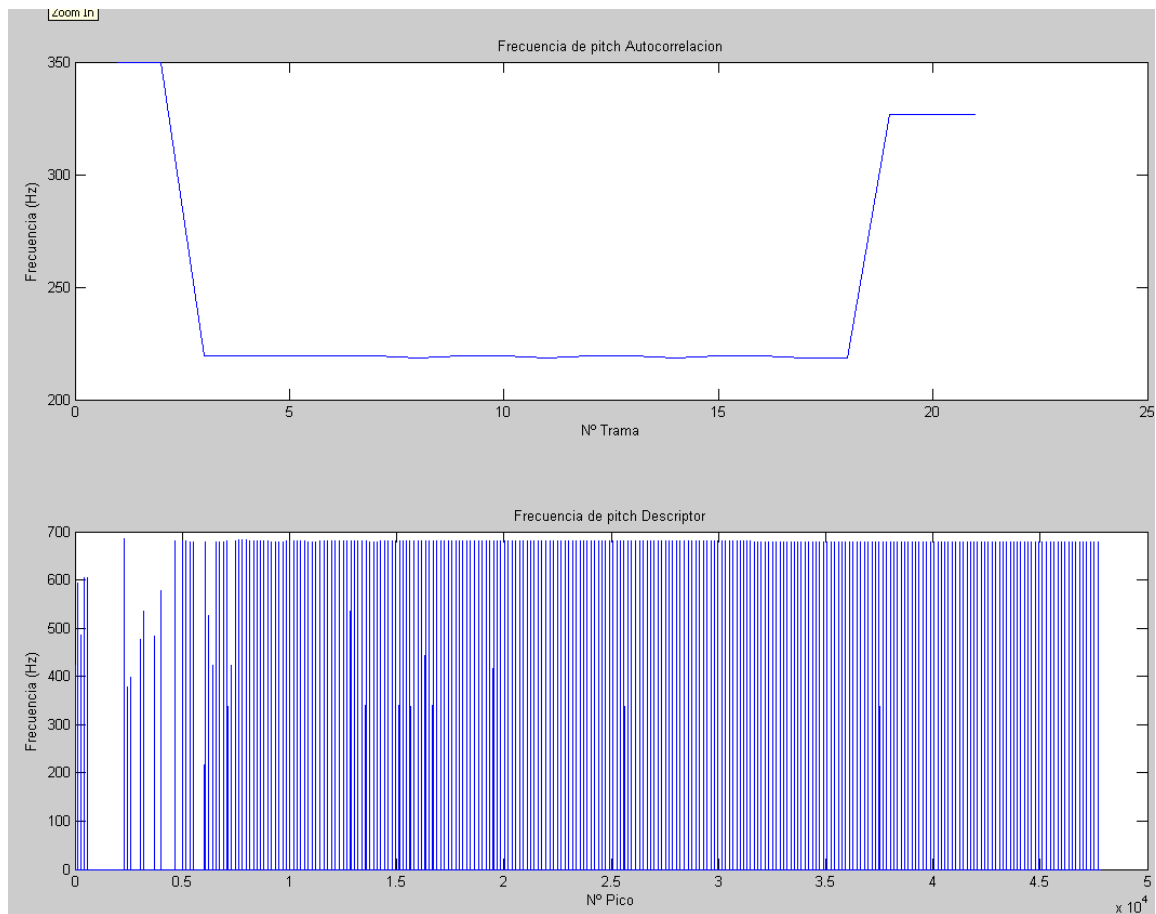


Figura 7.21: La figura superior es la frecuencia fundamental que se obtiene por medio de la autocorrelación. La figura inferior es la frecuencia fundamental que se obtiene con el método del descriptor.

En este caso se puede ver como se obtiene una frecuencia fundamental completamente distinta para ambos métodos.

En el caso de la autocorrelación se obtiene una frecuencia fundamental de unos 220 Hz, mientras que en el caso del descriptor se obtiene una frecuencia fundamental de 680 Hz, que se ha visto que es la correcta por el espectrograma de la figura 7.18.

Por tanto se puede concluir que para una señal con más de una fuente el método de autocorrelación no funciona, mientras que el método del descriptor funciona correctamente.

8. CONCLUSIONES Y LINEAS FUTURAS

8.1 Conclusiones:

1. El parámetro Bw/L (Ancho de Banda normalizado) permite discernir entre un pico de ruido y un pico perteneciente a un armónico.
2. El descriptor Bmf (Ancho de banda normalizado * frecuencia central de pico) permite agrupar los picos de los diferentes armónicos, pudiendo así obtener la frecuencia fundamental de la señal.
3. Para señales con una sola fuente tanto el método de autocorrelación como el método del descriptor funcionan correctamente y dan una frecuencia fundamental muy parecida.
4. Para señales con más de una fuente el método de autocorrelación falla, mientras que con el método del descriptor permite obtener la frecuencia fundamental del instrumento con frecuencia fundamental más baja.

8.2 Líneas Futuras de Investigación:

- La principal aportación de este proyecto es la detección de pitch de una señal polifónica, recordemos que se obtenía el pitch del instrumento de menor frecuencia fundamental, por tanto en futuros proyectos podrían obtenerse la frecuencia fundamental de cada uno de los instrumentos que aparecen en la señal.
- Por otra parte una vez obtenida la frecuencia fundamental de un instrumento se podrían obtener cada uno de sus armónicos y poder separar de este modo un solo instrumento de la señal polifónica.
- El problema del solapamiento de parciales es uno de los principales inconvenientes a la hora de realizar estimación multi-pitch de señales musicales ya que implica ambigüedad espectral. Esto se debe a que la música Occidental, basada en la armonía de los eventos existentes, se favorece la aparición de eventos con relación armónica entre sí. La técnica utilizada en este proyecto es una herramienta que ofrece un buen rendimiento en la estimación multi-pitch de señales polifónicas pero no ante señales de este tipo cuando gran parte de los parciales de los eventos se encuentran solapados en frecuencia. Por ello esta es la principal línea de investigación a seguir.

9. BIBLIOGRAFIA

Para la realización de este proyecto se utilizaron las siguientes referencias:

- [1] L. Cohen, “Time-Frequency Analysis: Theory and Applications”, Prentice Hall Signal Processing Series.
- [2] X. Serra “Musical Sound Modeling with Sinusoids plus Noise”, Computer Music Journal Bol. 14 Num 4
- [3] F. Teberio “Identificación de Sinusoidales no estacionarias para descomposición de audio y detección de pitch” Universidad de Navarra, 2009
- [4] M. Alonso, G. Richard, and B. David. “Extracting note onsets from musical Recordings”. ICME. Amsterdam, 2005.
- [5] J. W. Pitton, “Adapting Multitaper Spectrograms to Local Frequency Modulation”, IEEE 2000
- [6] F. Plante, G. Meyer & W. A. Ainsworth, “Improvement of Speech Spectrogram Accuracy by the Method of Reassignment”, IEEE 1998
- [7] N. Ray “Linear Methods for Classification”, Lecture notes for CMPUT 466/551.
- [8] F. J. Cañadas “Investigación y desarrollo de técnicas de estimación multi-pitch y su aplicación a la transcripción automática de señales musicales polifónicas”, Universidad de Jaén, 2009

Proyecto fin de carrera:

ESTUDIO DE PITCH EN SEÑALES POLIFÓNICAS

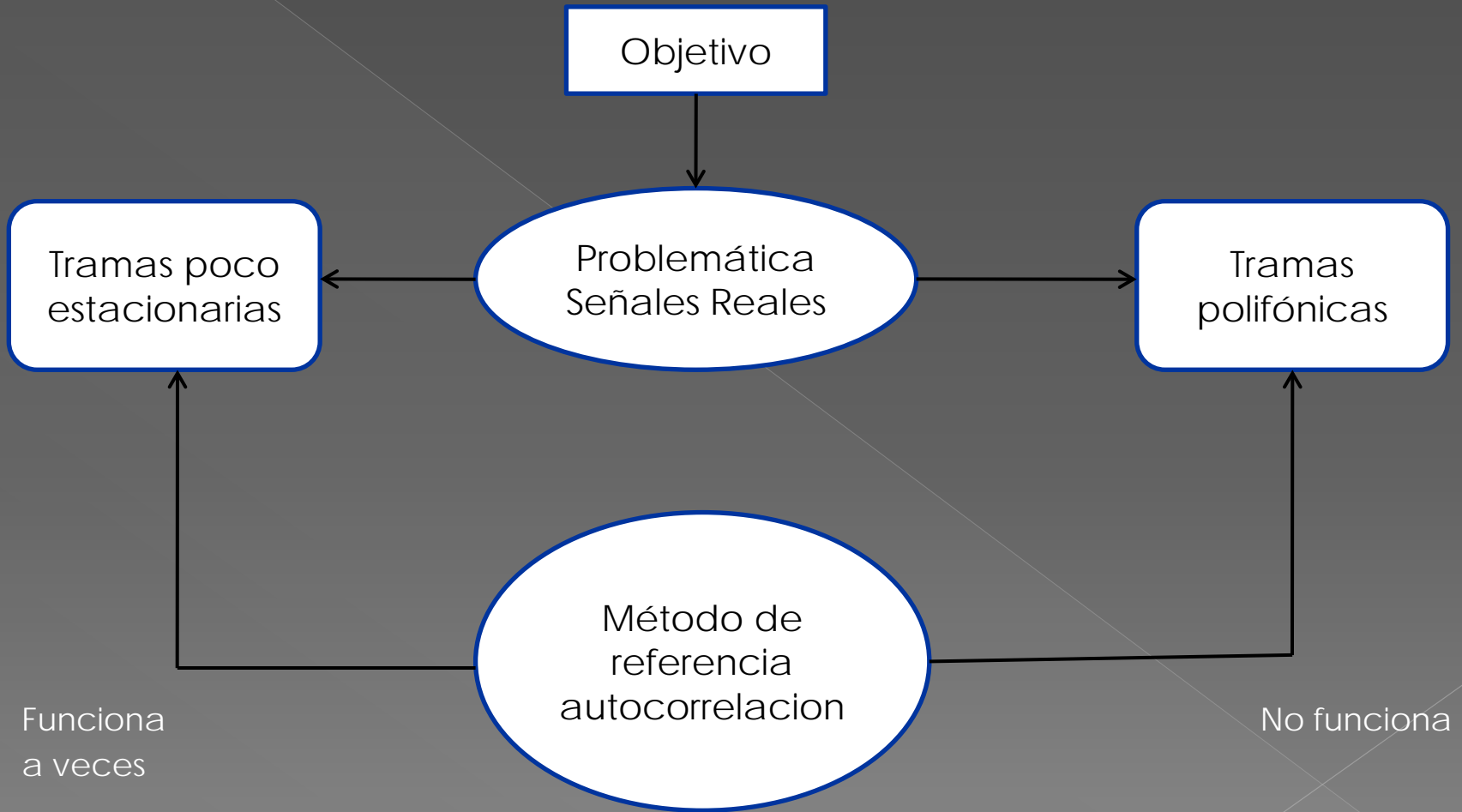
Autor:

**Raúl Zalba Oteiza
Miroslav Zivanovic**

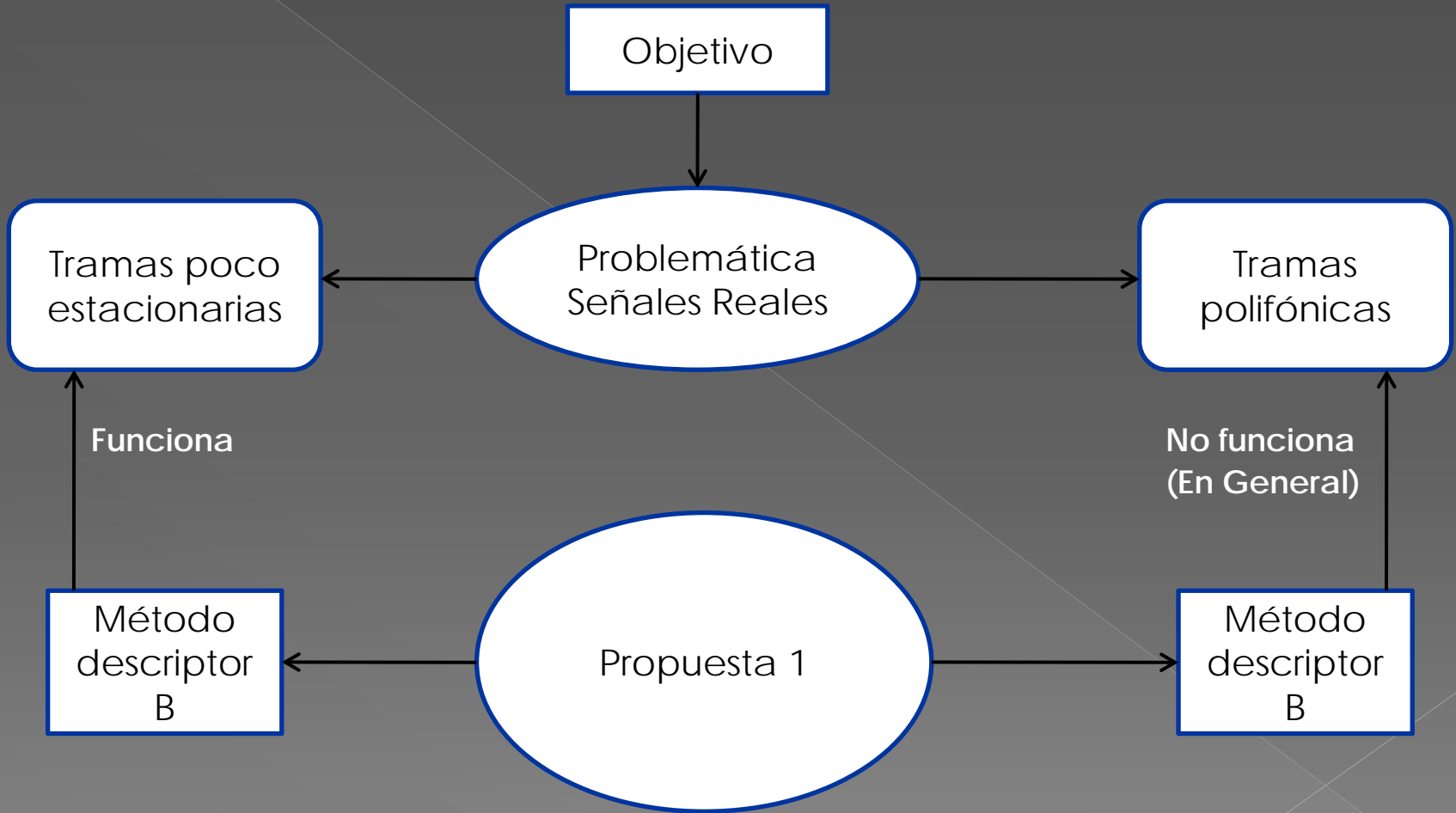
Objetivo del proyecto

- El objetivo principal de este proyecto es el calculo del pitch de una fuente de una señal polifónica.
- Para ello utilizaremos un método al que denominaremos método del descriptor y lo compararemos con el método de autocorrelación ya existente.

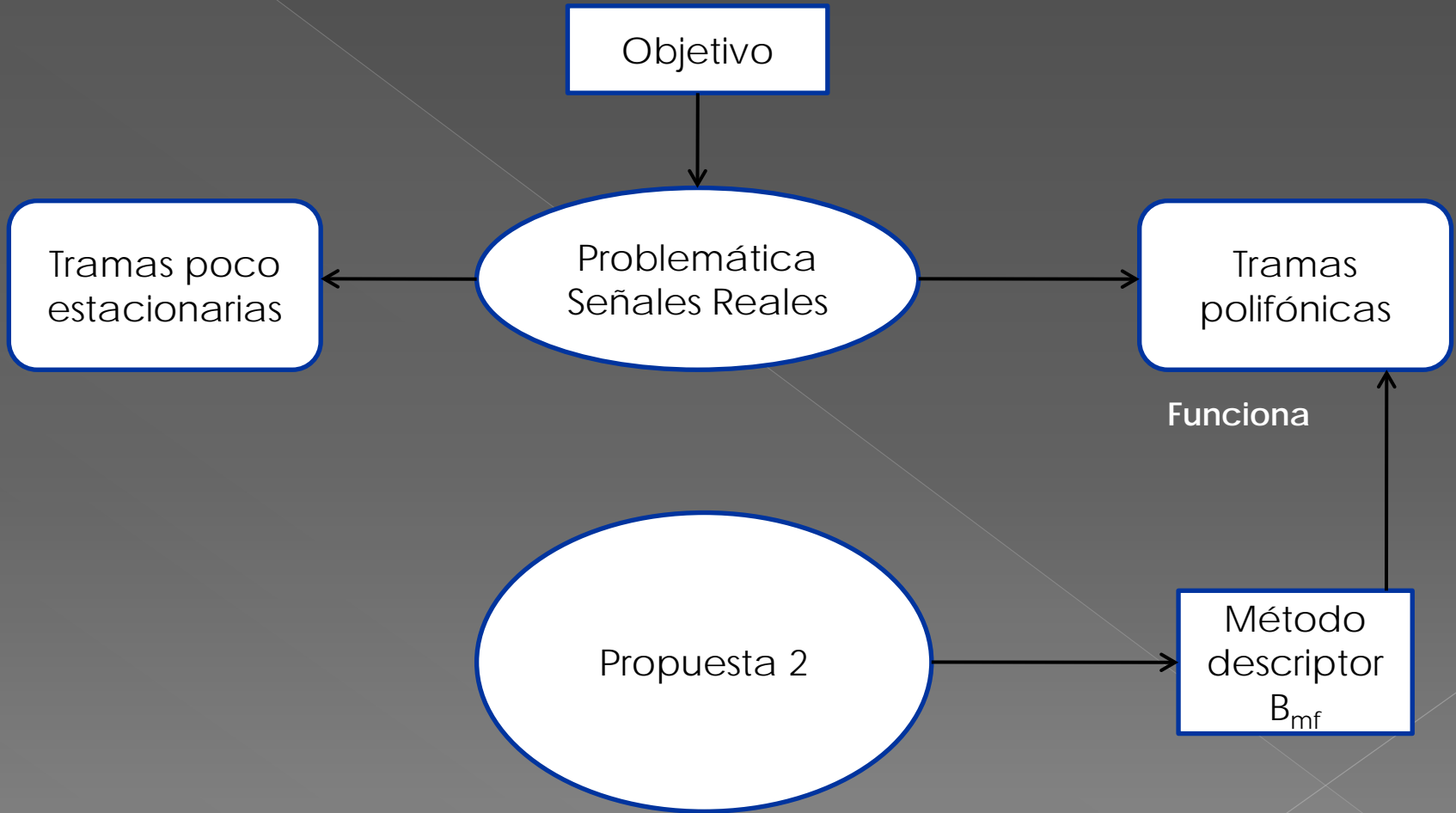
Problemática



Problemática



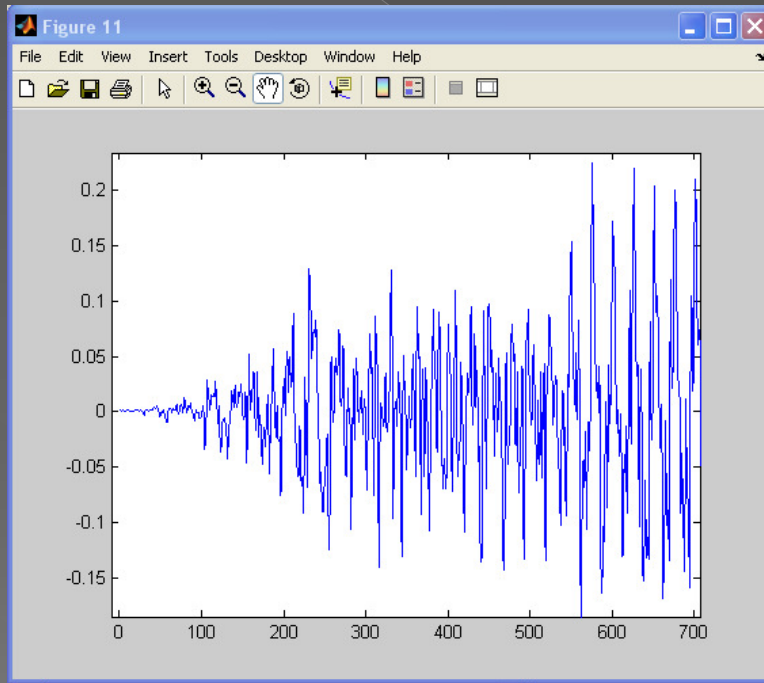
Problemática



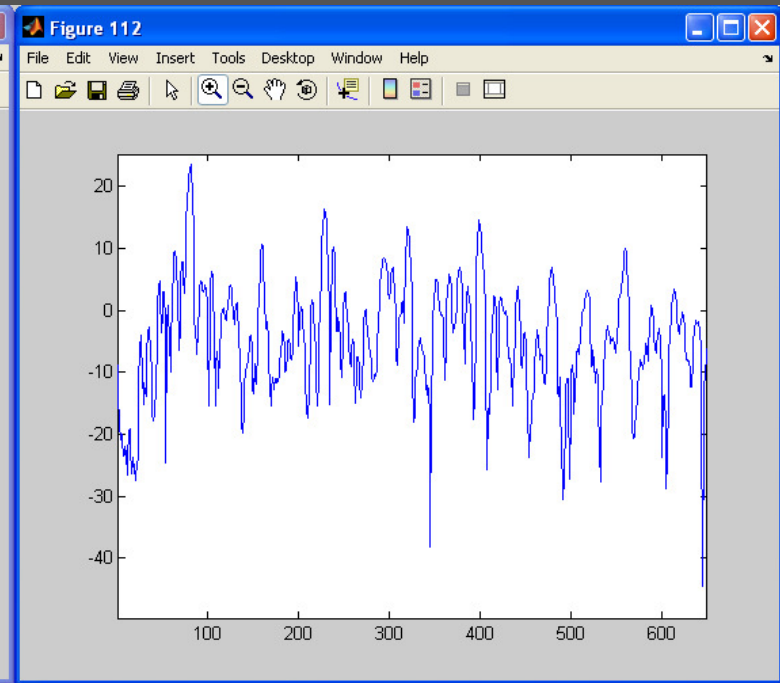
B_{mf} = Ancho de banda normalizado(B) x frecuencia central de pico espectral

Problemática

- Tramas poco estacionaria:



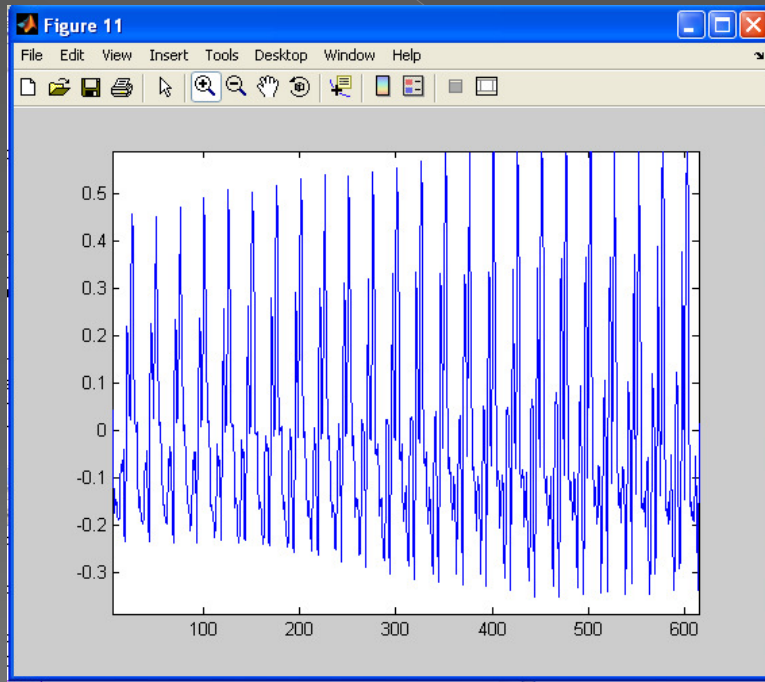
Dominio tiempo



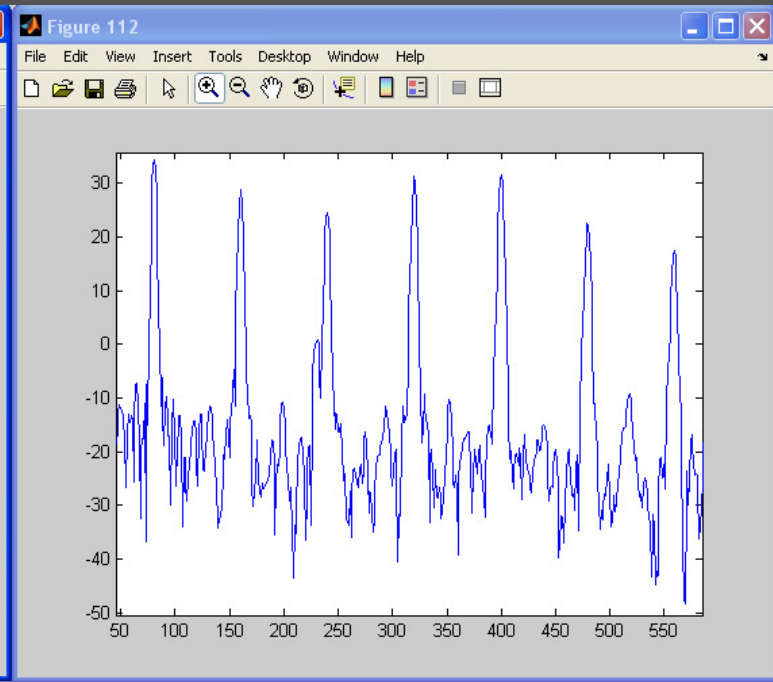
Dominio Frecuencia

Problemática

- Trama cuasi-estacionaria:



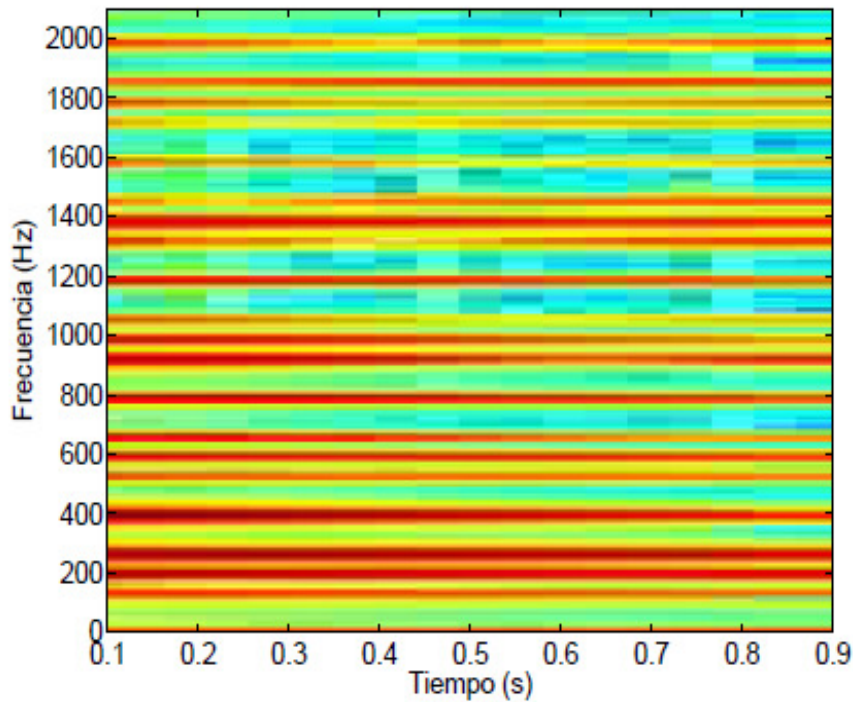
Dominio tiempo



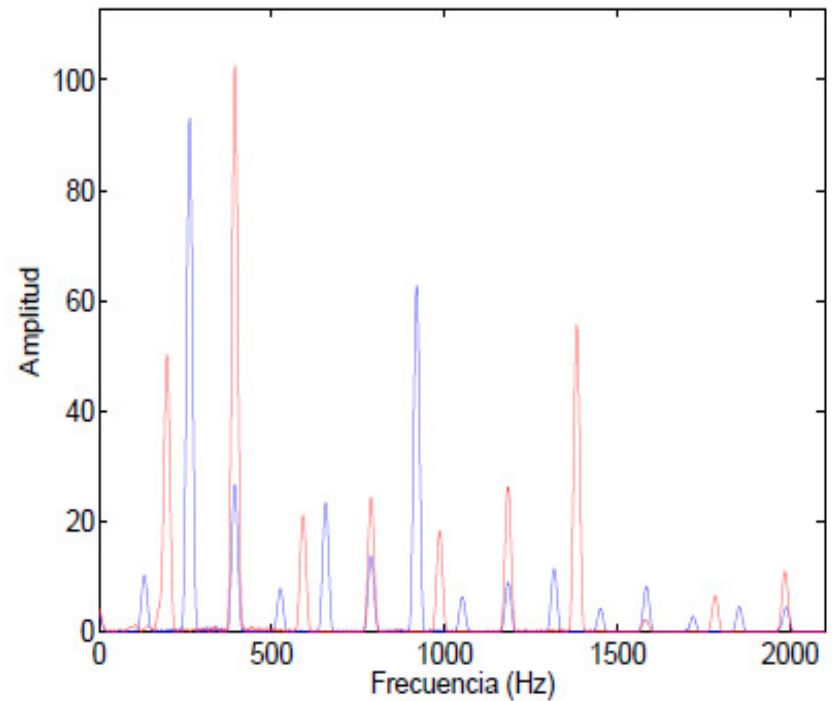
Dominio Frecuencia

Problemática

- Solapamiento de frecuencias debido a varias Fuentes:

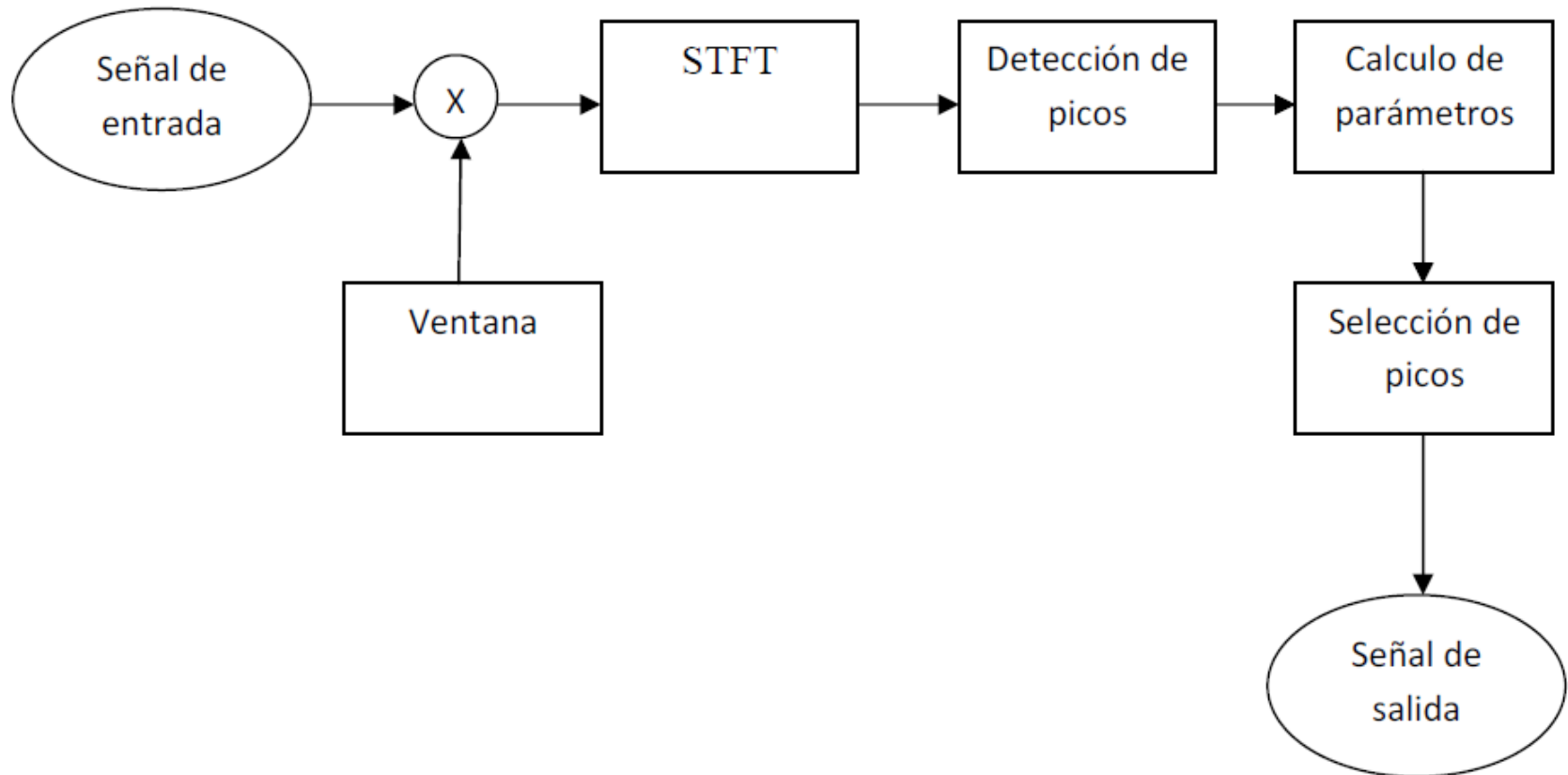


Espectrograma



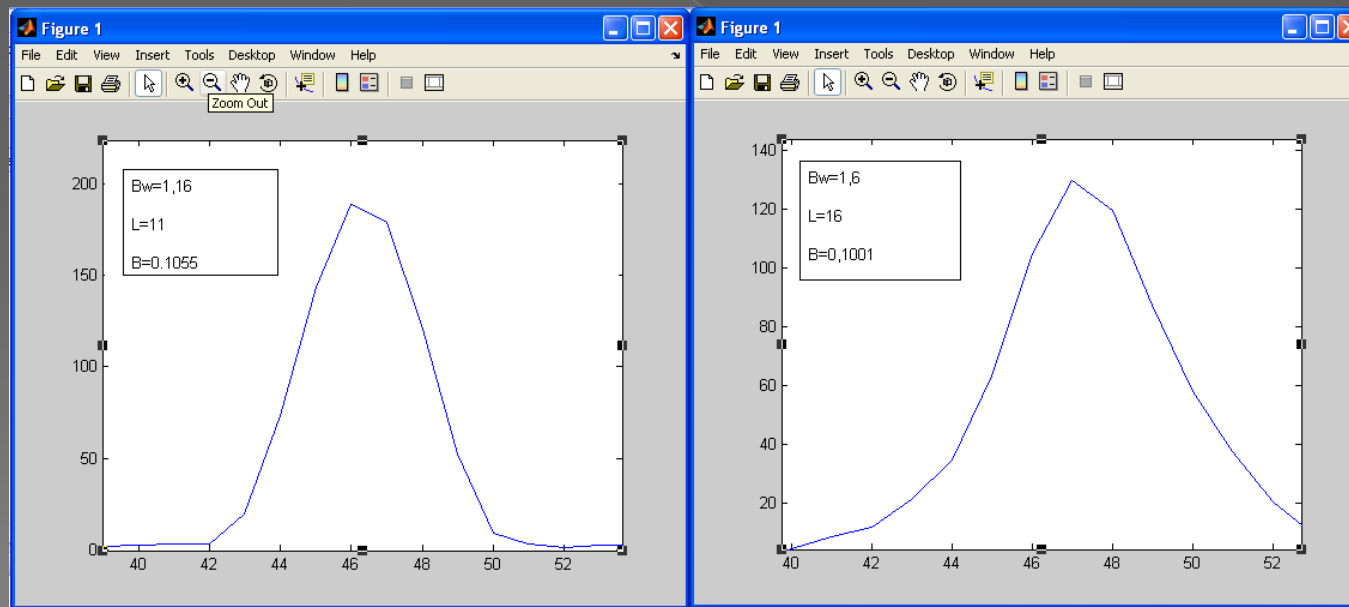
Dominio Frecuencial

Esquema General:



Clasificación de picos espectrales- Descriptor B

- Por cada pico, calcularemos 3 parámetros:
- Ancho de Banda de pico(B_w)
- Ancho de Banda Absoluto de pico(L).
- Frecuencia central del pico(m_f)
- Descriptor $B = B_w/L$ (Ancho de Banda Normalizado)



Tono puro

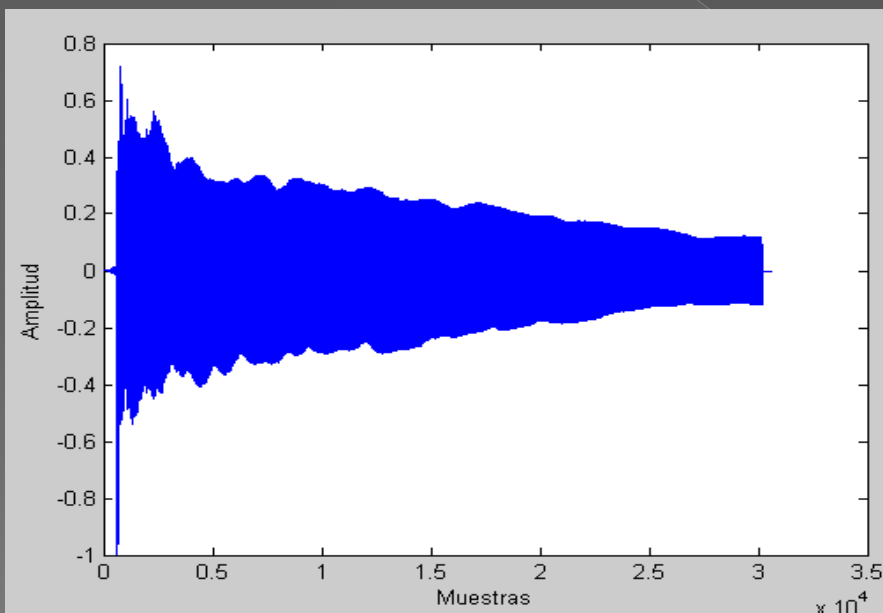
Señal modulada

Clasificación de picos espectrales-Descriptor B_{mf}

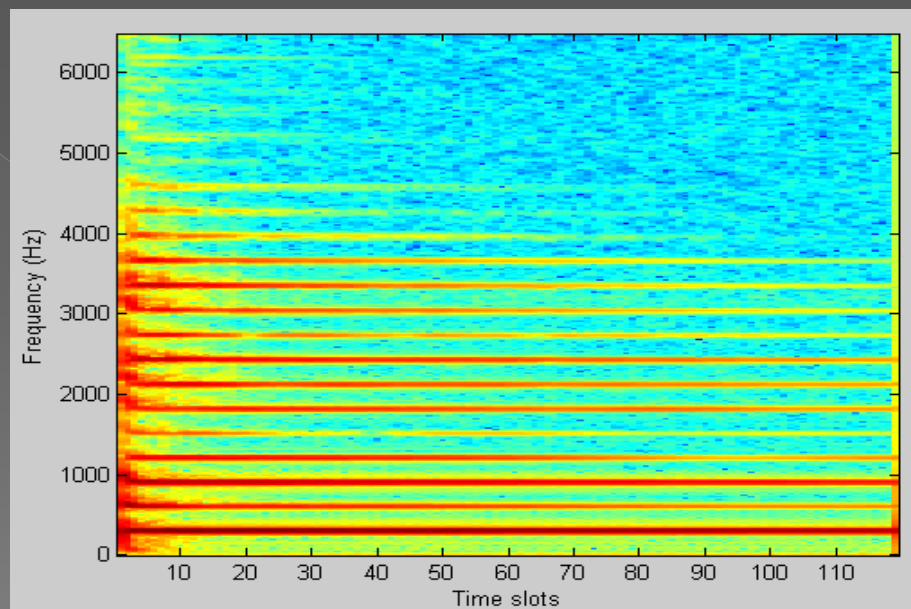
- ◉ Este descriptor es simplemente $B(\text{Ancho de banda normalizado}) \times mf$ (frecuencia central del pico)
- ◉ Con los valores de este descriptor podemos agrupar los diferentes armónicos de la señal.
- ◉ Nos permite seleccionar los picos pertenecientes a la Frecuencia fundamental.

RESULTADOS EXPERIMENTALES

- Señal con una sola fuente
- Nota de una guitarra.



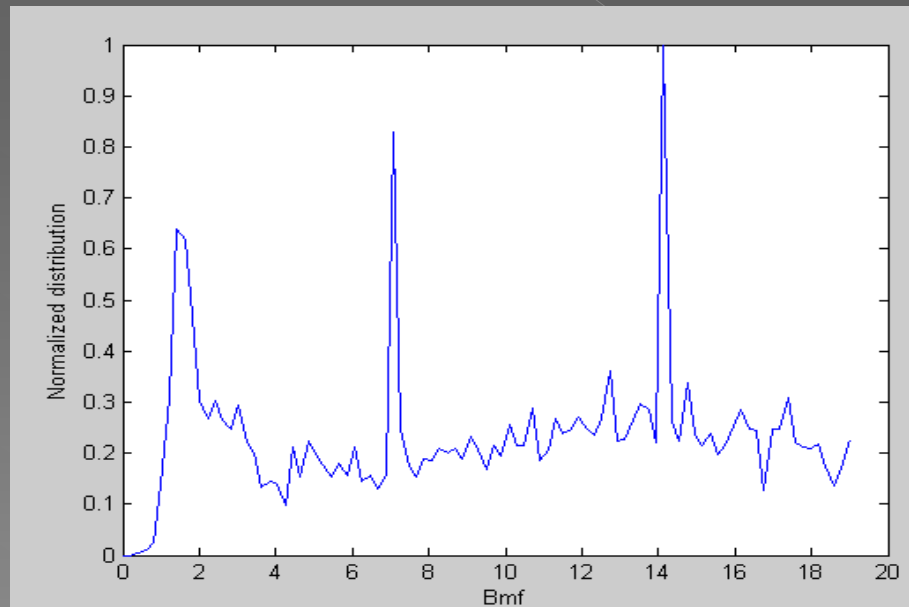
Forma de la señal



Espectrograma

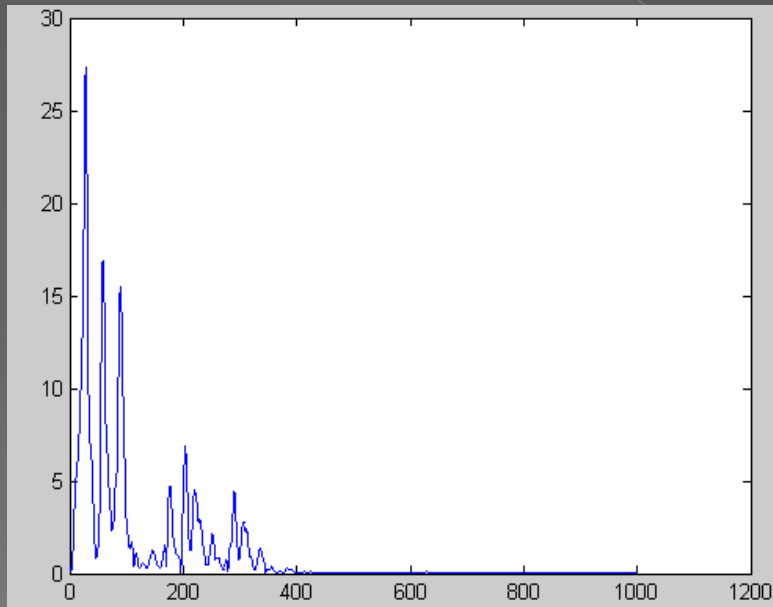
RESULTADOS EXPERIMENTALES

- Se divide la señal en tramas, a través de una ventana.
- Calculamos la STFT de cada trama.
- Buscamos todos los picos espectrales de la trama.
- Calculamos los parámetros de cada pico.
- Con los valores del descriptor B_{mf} creamos este histograma:

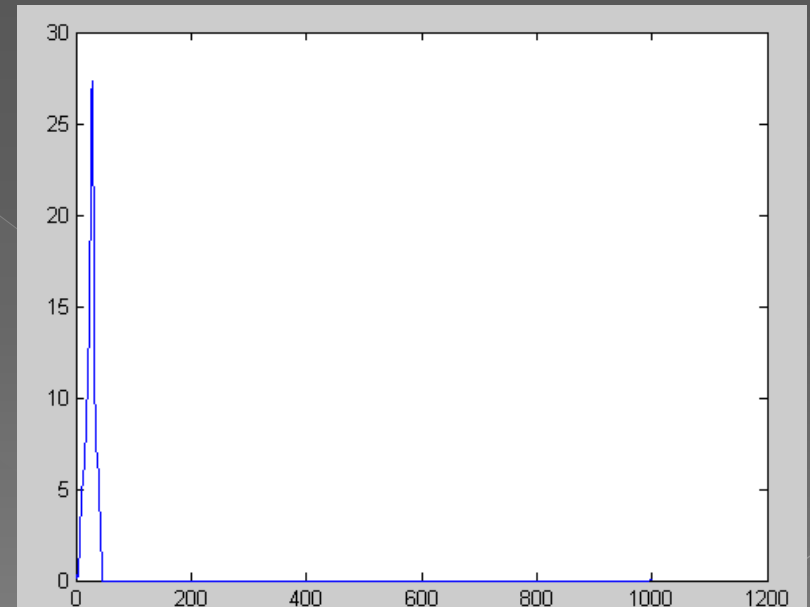


RESULTADOS EXPERIMENTALES

- Espectros de una trama.



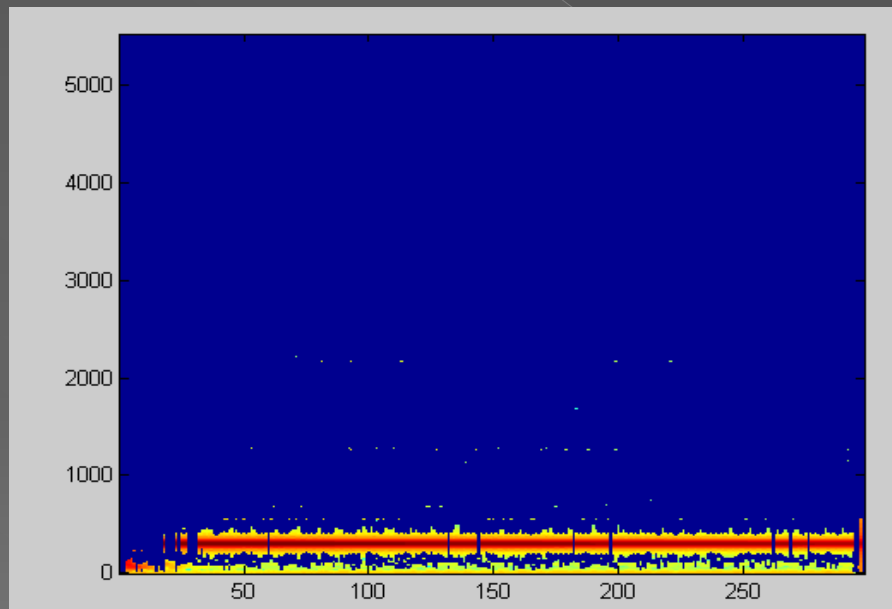
Espectro inicial



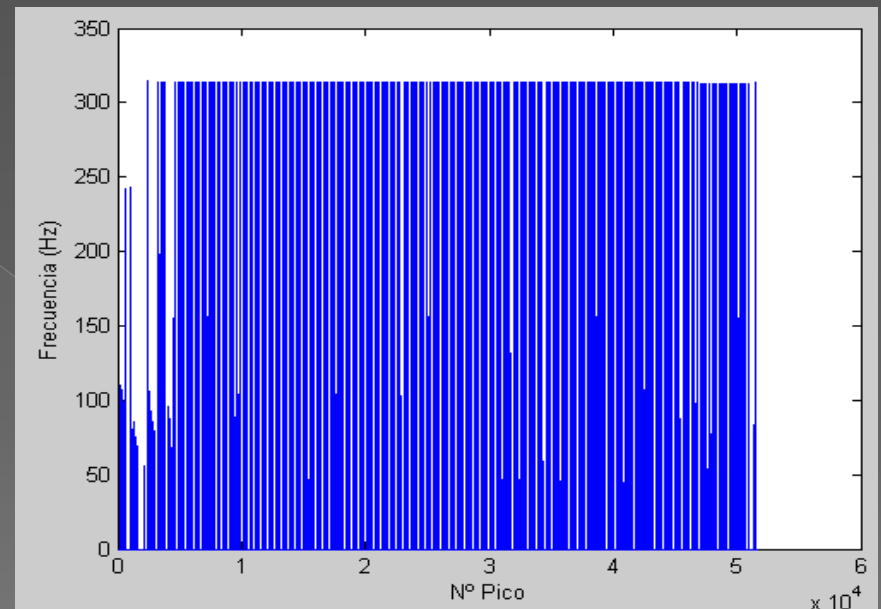
Espectro modificado

RESULTADOS EXPERIMENTALES

- Frecuencia Fundamental



Espectrograma

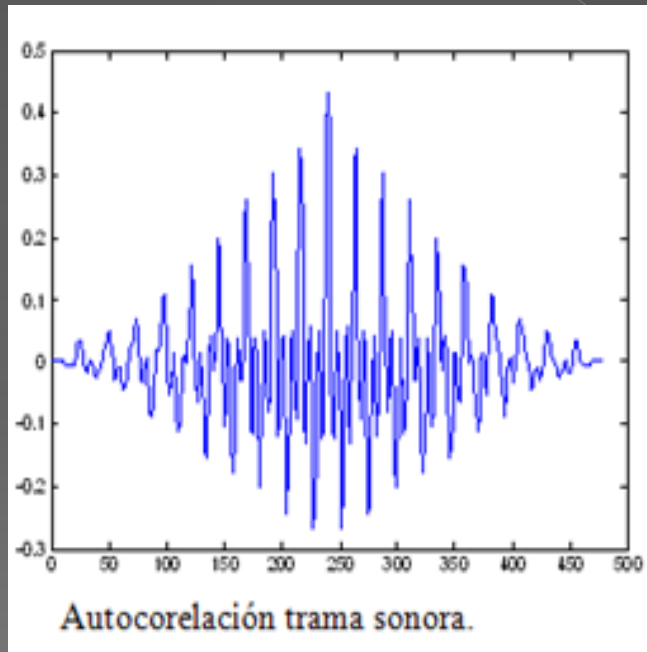


Frecuencia central pico (mf)

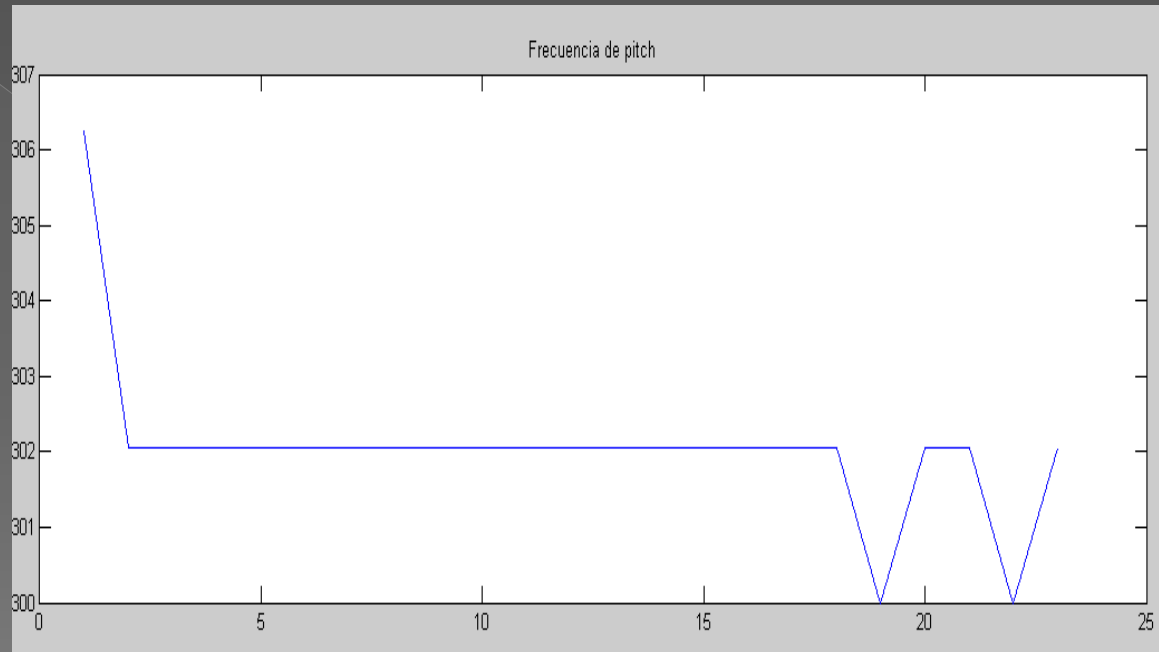
- Mejora representación con post-procesado.

RESULTADOS EXPERIMENTALES

● Método Autocorrelación



Ej. Trama

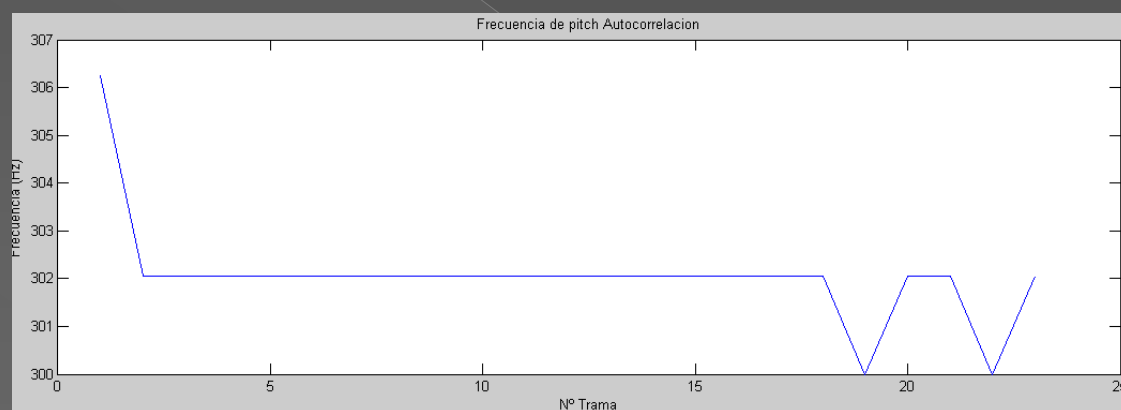


Frecuencia fundamental

RESULTADOS EXPERIMENTALES

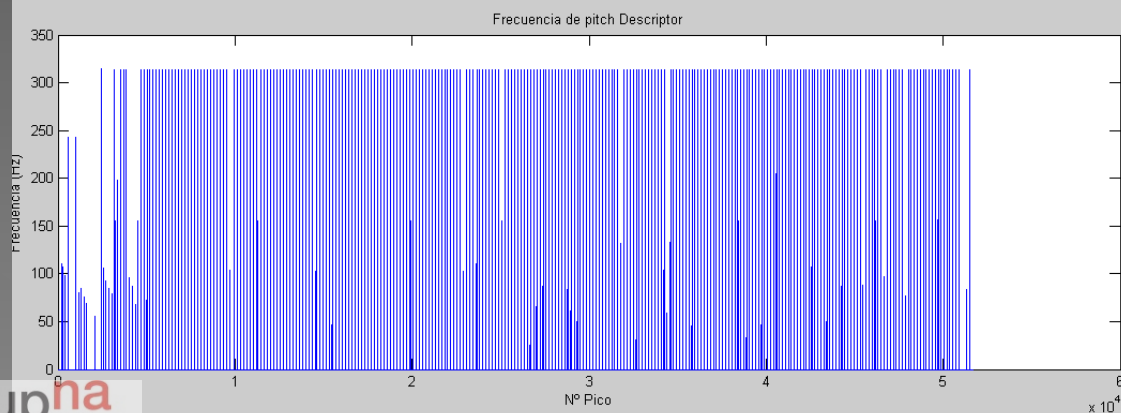
○ Comparación de métodos

Fpitch= 310 Hz



Autocorrelación

Fpitch=302 Hz

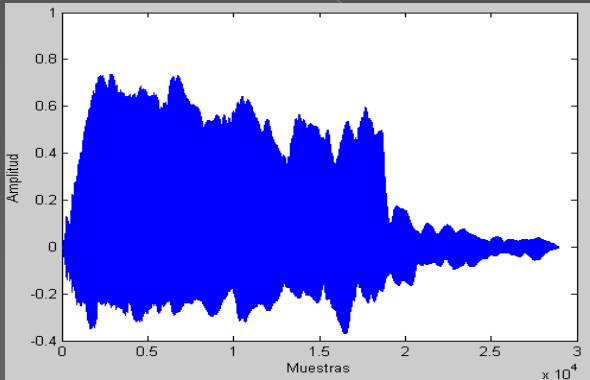


Método descriptor

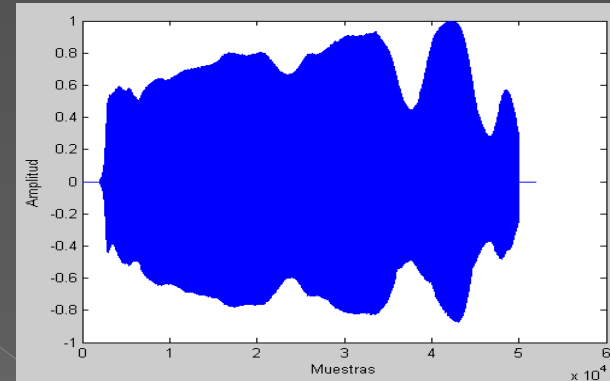
Fpitch=310 Hz

RESULTADOS EXPERIMENTALES

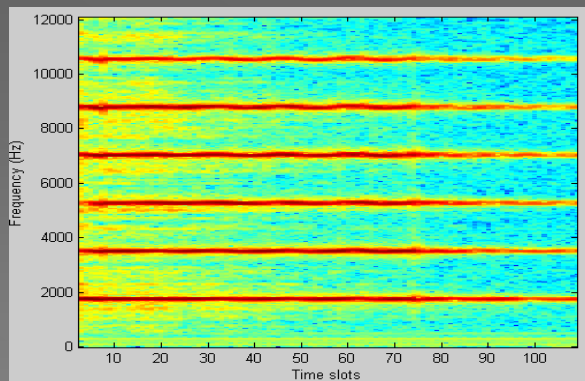
- Señal con dos fuentes (piano y oboe)



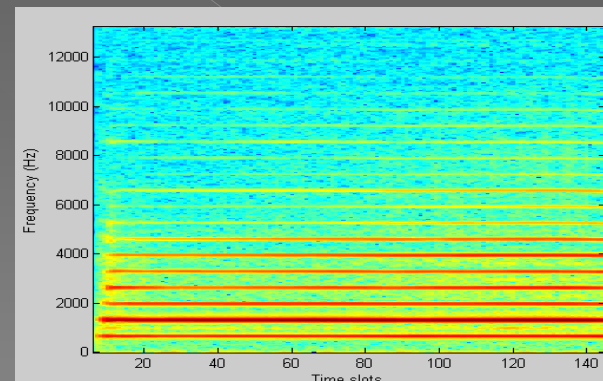
Forma de señal de piano



Forma de señal de oboe



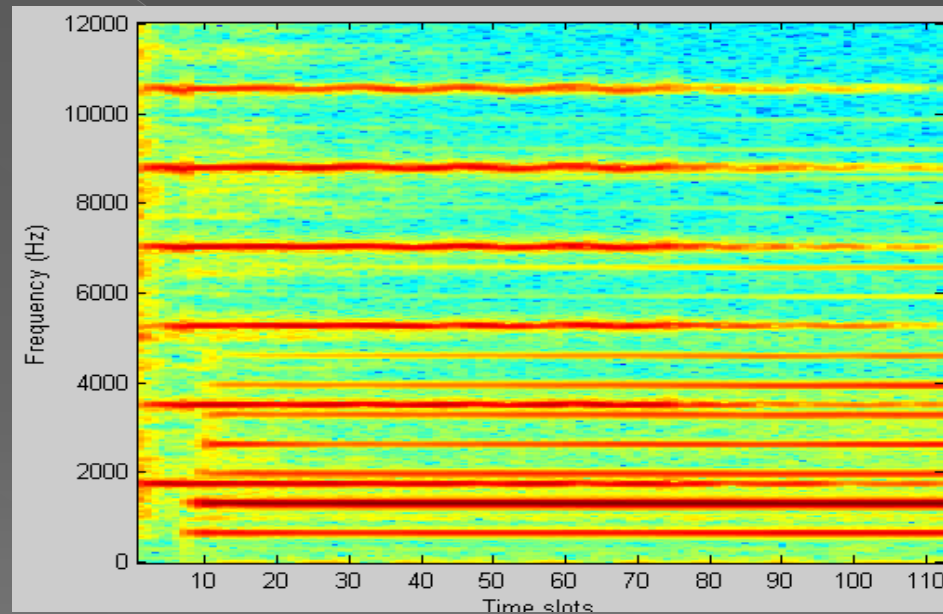
Espectrograma piano



Espectrograma oboe

RESULTADOS EXPERIMENTALES

- Suma de señales.

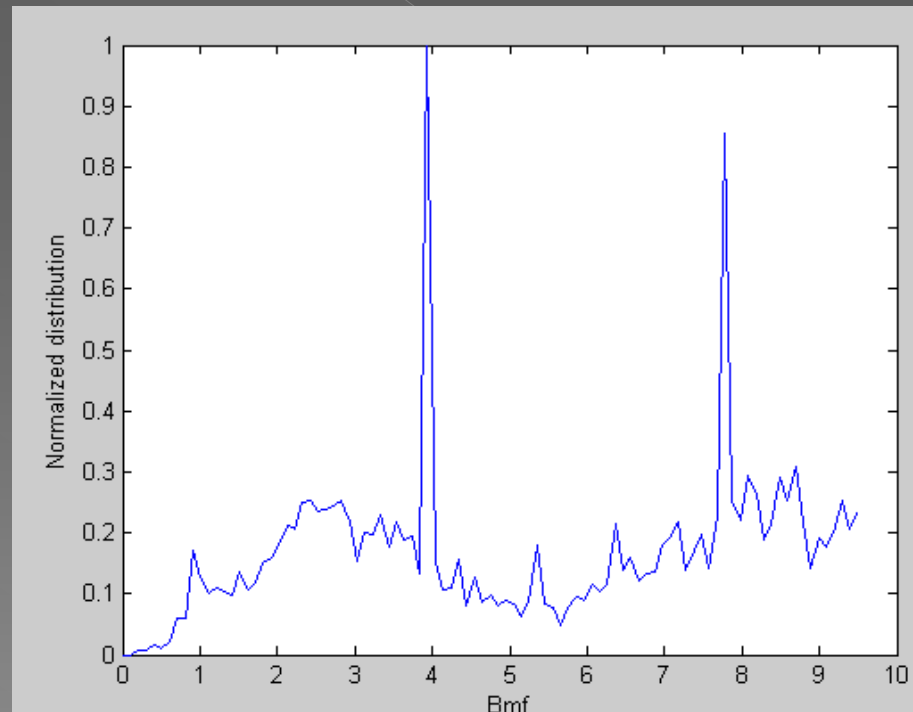


Espectrograma

- Diferenciación de armónicos

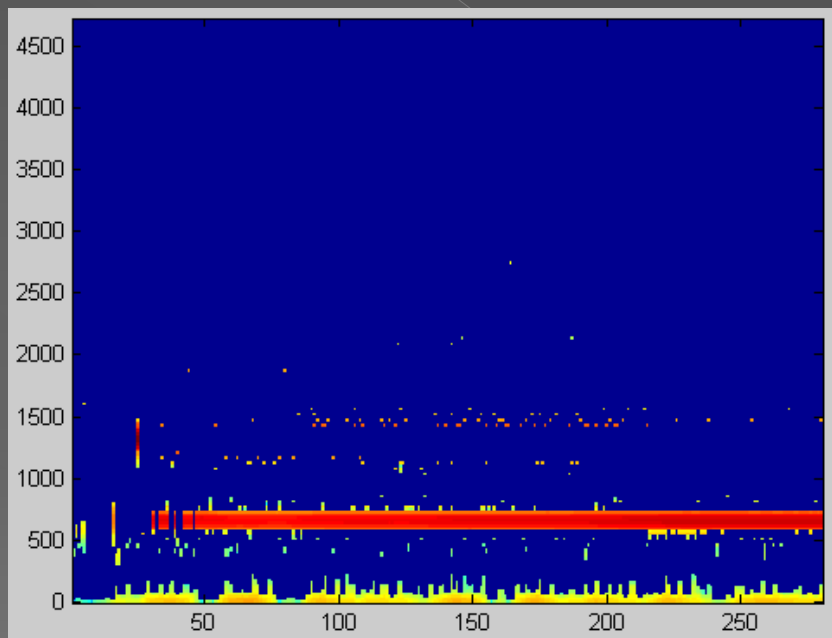
RESULTADOS EXPERIMENTALES

- Se divide la señal en tramas, a través de una ventana.
- Calculamos la STFT de cada trama.
- Buscamos todos los picos espectrales de la trama.
- Calculamos los parámetros de cada pico.
- Histograma con los valores del descriptor B_{mf} :

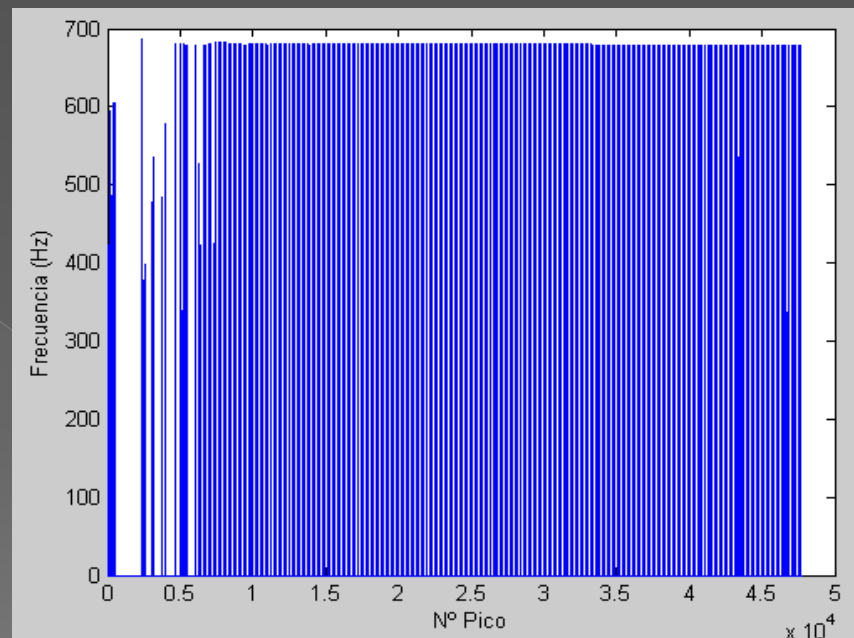


RESULTADOS EXPERIMENTALES

● Frecuencia Fundamental



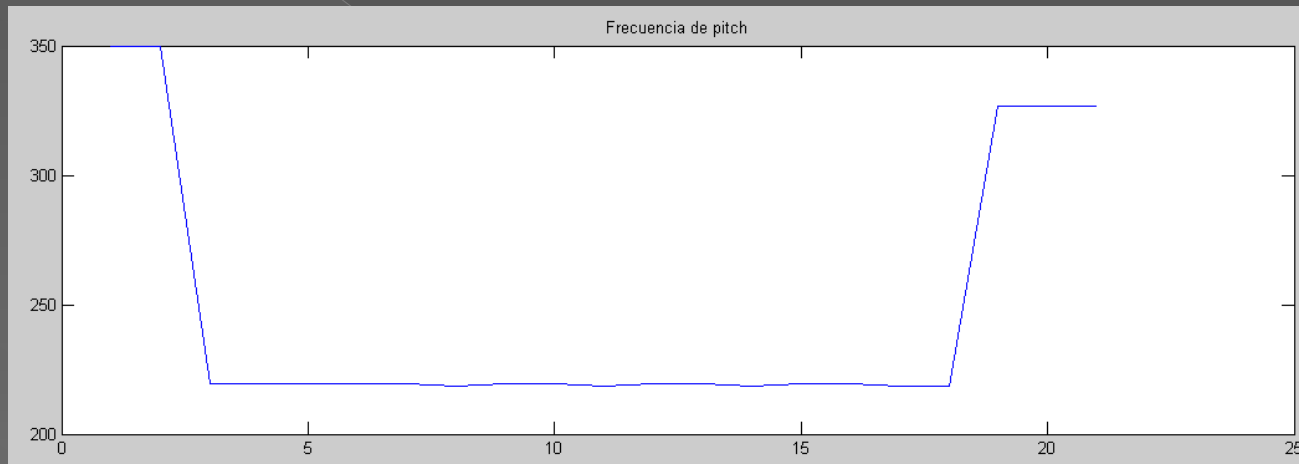
Espectrograma



Frecuencia central pico (mf)

RESULTADOS EXPERIMENTALES

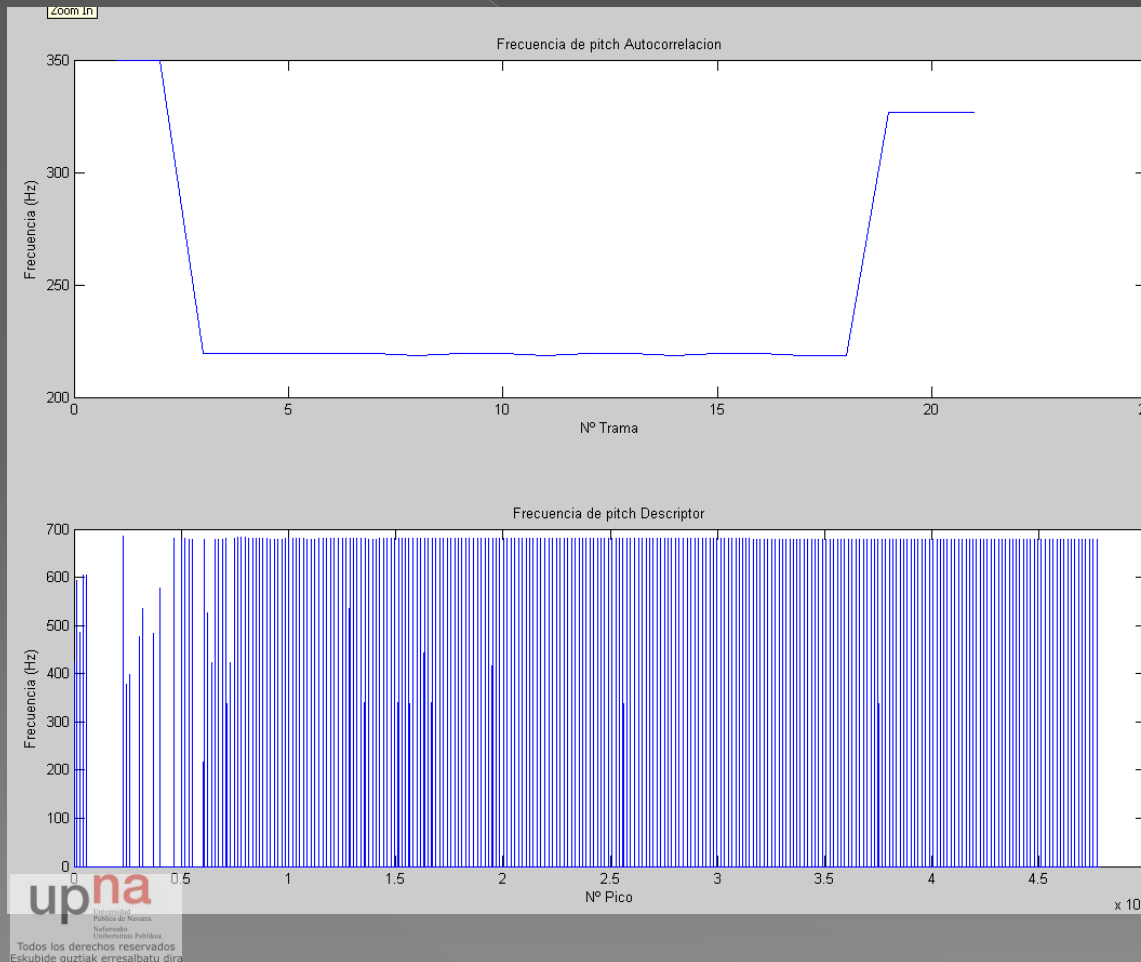
● Método Autocorrelación



Frecuencia Fundamental

RESULTADOS EXPERIMENTALES

○ Comparación de métodos



Fpitch= 680 Hz

Autocorrelación

Fpitch=250 Hz

Método descriptor

Fpitch=680 Hz

Conclusiones

- ◉ El descriptor $B=B_w/L$ (Ancho de Banda normalizado) permite discernir entre un pico de ruido y un pico perteneciente a un armónico.
- ◉ El descriptor $B_{mf} = B \times mf$ (Ancho de banda normalizado x frecuencia central de pico) permite agrupar los picos de los diferentes armónicos, pudiendo así obtener la frecuencia fundamental de la señal.
- ◉ Para señales con una sola fuente tanto el método de autocorrelación como el método del descriptor funcionan correctamente y dan una frecuencia fundamental muy parecida.
- ◉ Para señales con más de una fuente el método de autocorrelación falla, mientras que el método del descriptor permite obtener la frecuencia fundamental de la fuente con frecuencia fundamental más baja.

Líneas Futuras de investigación:

- La principal línea de investigación a seguir sigue siendo el problema de solapamiento de parciales.
- Una vez obtenida la frecuencia fundamental de una fuente se podrían obtener cada uno de sus armónicos y poder separar de este modo esa fuente de la señal polifónica.
- Las señales de voz introducen mas ruido que los instrumentos, lo que dificulta el calculo de su frecuencia fundamental, este seria otra línea a investigar.

FIN

GRACIAS MIRO !!!!!!!